UNDERSTANDING THE MECHANISMS OF THE PHOTOPERIOD FLOWERING PATHWAY IN SOYBEAN

BY

WILLIAM B. PRICE

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Master's Committee:

Assistant Professor Yoshie Hanzawa
Professor Lila Vodkin
Professor Randall Nelson
Associate Professor Matthew Hudson

ABSTRACT

Flower transition is the shift from the vegetative phase to the reproductive phase of a plant. Flowering transition determines a plant's survivability and reproductive success. Flowering transition is triggered by environmental and endogenous cues. The general purpose of this thesis was to gain insight into the genetics of flower transition in the major crop specie, soybean (*Glycine max*).

Chapter One introduces the long day dicot model species *Arabidopsis thaliana* that serves as the paradigm of understanding the genetic basis of flower transition. The six flowering pathways of *A. thaliana* are described in detail to introduce the reader to the complex molecular networks that act to initiate flowering. Next, the focus narrows to advance the understanding of the photoperiodic flowering pathway. The mechanisms of photoperiod flowering in short day monocot model species *Oryza sativa* are introduced and described to illustrate conservation and divergence between the photoperiod flowering pathway components in the two model species. Finally, the known components of soybean photoperiod flowering are introduced, and the question is raised whether *CONSTANS* a known flowering promoter in *Arabidopsis* and rice is conserved in soybean.

Chapter Two discusses the research to characterize *CO* homolog expression in soybean across two day length treatments by qRT-PCR through a candidate gene approach. The candidate gene approach applies bioinformatics techniques along with the recently published *G. max* genome to infer evolutionary history of the *CO* gene to identify homologs in the soybean genome for further study. Once identified, the genes can then be characterized via expression analysis to determine possible function analogous to *CO*.

Chapter Three discusses the research to characterize *CO* homolog expression in soybean across two day length treatments by RNA sequencing to support the qRT-PCR results and to characterize differences in expression between seven soybean genotypes. Although the results do not provide definitive results for differences across genotypes, they do provide two putative candidates for *CO*-like genes that function in photoperiodic flowering pathway to induce floral initiation.

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 Flowering

In an angiosperm, floral initiation is a unique and major physiological change from vegetative growth to reproductive development. Flowering is essential to the plant's survivability and reproductive success. In agricultural development, flowering is an important target of domestication. By manipulating flowering, it is possible to diversify cultivation strategies to optimize yield in specific locations. For example, variation in flowering time is selected for in soybean. Since day length changes as locations move further and further from the equator soybean maturity groups are classifications that indicate latitudes suited for individual lines. Similarly, in wheat, the manipulation of vernalization has resulted in optimal yields based on the varying season lengths and temperatures. Winter and spring cultivars have allowed for double cropping of certain lands.

Highly sensitive and complex regulatory networks control the timing of flowering in response to internal cues and external environmental fluctuations. Flowering gene networks are well studied in model species, however, not in important crop species. Therefore, there is much to be known on the molecular basis of flowering for crop species.

1.2 Flowering Pathways of Long Day Model *Arabidopsis thaliana*

Transition to flowering is an important ecological trait in *Arabidopsis thaliana* in response to various endogenous and environmental cues, such as photoperiod, light quality, ambient temperature, and vernalization, as well as biotic and abiotic stresses. Long day flowering plant *Arabidopsis thaliana* is a model for understanding a complex genetic network by forward and reverse genetic approaches (Srikanth and Schmid, 2011). In *Arabidopsis thaliana*, there are six major genetic pathways that control flowering: photoperiod, vernalization, autonomous, gibberellin (GA), ambient temperature, and age pathway. These pathways control a shift from vegetative development to reproductive development by regulating floral meristem identity genes in the shoot apical meristem (SAM), which results in flower organ development.

Environmental Signals of Floral Initiation

Plants are classified as long day, short day, or day-neutral, based on their flowering response to day lengths. Long day (LD) plants are characterized as plants that initiate flowering when the day length exceeds a particular threshold. Short day (SD) plants initiate flowering when the length of night exceeds a certain threshold once the plant has matured. Day-neutral plants initiate flowering due to factors independent of day length (Garner and Allard, 1920, 1923).

In the last century, many hypotheses were developed and tested to explain how plants perceive photoperiod. Such theories involve an idea that plants require internal oscillators to synchronize endogenous rhythms with exogenous signals

(Bünning, 1960; Pittendrigh and Minis, 1964). Plants perceive light in the leaves, but not the shoot apical meristem, suggesting that flowering is induced by an internal factor that is transported from the leaf into the SAM under appropriate photo-inductive conditions (Zeevart, 2006).

Light quality and quantity, not only day length, also play key roles in floral initiation. Plants are susceptible to shade avoidance syndrome, which results in precocious flowering in response to a shift in the ratio of red to far-red incoming light (Smith and Whitelam, 1997). Additionally, blue light quality at high intensity disrupts normal flowering patterns (Tsughiya and Ishiguri, 1981). However important, light is not the lone regulator involved in floral transition.

Temperature is another external factor that plants have developed mechanisms for identifying and controlling different life processes, like flowering. For example, many *Arabidopsis thaliana* ecotypes require vernalization, a period of cold treatment (Johanson et al., 2000). Over time individuals observed that the difference between winter and spring annuals was a result of an obligate vernalization requirement in winter annuals (Chouard, 1960). A distinction can be made between prolonged exposure to cold and ambient temperature. By a separate mechanism, flowering times can be altered due to an increase in ambient temperature (Lee et al., 2007).

Internal Signals of Floral Initiation

Plants require many internal, or developmental, factors to promote flowering in conjunction with external, or environmental, cues. Gibberellic acid (GA) is capable of inducing flowering during non-inductive photoperiods and without vernalization

(Langridge et al., 1957; Chandler et al., 1994). In long day growth conditions, GAs have little influence on flowering time, but the GA pathway in short day growth conditions becomes the major player in initiating flowering, where the plant is deprived of photoperiod induced flowering (reviewed by Mutasa-Gottgens and Hedden, 2009). However, GA is not universal in inducing flowering, and, therefore, is not a true flowering hormone (Zeevart 1983). In some species, GAs have little effect or inhibit floral initiation (Zeevart 1983). Nevertheless, GA plays a significant role in the control of flowering in many species.

The cumulative effect and ultimate redundancy of these pathways in controlling flowering is predicated on many early observations of a diverse array of plants without the application of modern plant genetics and molecular biology. However, since the advent of modern genetics the use of *Arabidopsis thaliana* as a model plant due to its small genome, out-crossing ability, large seed count, and short life cycle has illuminated a much deeper understanding of plant life processes, such as flowering, at the molecular level (Laibach et al., 1943). The resulting paradigm can serve as the foundation for the study of more complex and useful plant species.

Temperature-dependent and Autonomous Pathways

Vernalization is defined as the acquisition or acceleration of the ability to flower by a chilling treatment (Chouard, 1960) (Figure 1). Vernalization, in general, requires <7°C for a consistent period of 1-3 months (Chouard, 1960; Purvis and Gregory, 1952). *FRIGIDA* (*FRI*) is a key gene that controls vernalization (Johanson et al. 2000). Early-flowering ecotypes from low latitudes have loss-of-function mutations in the *FRI* gene (Johanson et al. 2000). *FRI* is a positive regulator of

FLOWERING LOCUS C (*FLC*), a known MADS box transcription factor and repressor of flowering (Michaels et al. 1999). *FLC* antisense transcripts, *COOLAIR* (cold induced long antisense intragenic RNA) resulting from cold induction, act to transiently silence *FLC* transcription, and subsequent recruitment of a PHD-PRC2 complex (VERNALIZATION INSENSITIVE 3 (VIN3) – POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) that includes VERNALIZATION 2 (VRN2)) bestows epigenetic memory, "memory of winter" (Swiezewski, 2009; De Lucia et al., 2008; Baürle et al., 2006). Thus, plants are flowering-competent once vernalization occurs.

The autonomous and vernalization pathways are interconnected as they both regulate *FLC* expression. The autonomous pathway compiles a multitude of genetic factors involved in RNA processing and epigenetic regulation that down-regulate the floral repressor *FLC* (Simpson, G.G. 2004) (Figure 2). A series of proteins are involved in the pathway. FCA, FPA, and FLK are involved in RNA binding, FY in RNA processing, and FVE and FLD in chromatin regulation (Baürle et al., 2006). Recently, Baürle et al. proposed that the autonomous pathway is part of a widely conserved transcriptome surveillance mechanism and *FLC* is especially sensitive target due to selection for flowering time variation (Baürle et al., 2008).

In contrast, the pathway that perceives ambient temperature is suppressed by *FLC*, and depends on the gibberellin hormone (Balasubramanian et al. 2006). A moderate increase in growth temperature, from 23°C to 27°C, was sufficient to induce flowering comparable to the effect of long day light conditions (Lempe et al. 2005). There is a great deal of natural variation in temperature response in plants (Blazquez et al. 2003). Genes involved with alternative splicing were functionally

temperature-dependent. Specifically, alternative splicing of *FLOWERING LOCUS M* (*FLM*), a MADS box gene like *FLC*, was masked at higher temperatures suggesting that splicing is an important regulator of flowering (Balasubramanian et al. 2006).

Recently, a study uncovered that the protein kinase CK2 antagonizes the key circadian clock regulator CIRCADIAN CLOCK-ASSOCIATED 1 (CCA1), which binding affinity increases at higher temperatures, at higher temperatures to assure the circadian clock does not run "too fast" (Portoles and Mas, 2010). This study is an interesting example of the power of temperature cues, and the plant's ability to evolve coping mechanisms to ensure overall fitness.

Lee et al. (2007) propose that ambient temperature signaling is mediated by MADS box protein SHORT VEGETATIVE PHASE (SVP). SVP binds to the CArG motifs on *FT* and *SOC1* promoters repressing floral induction (Hartmann et al., 2000). SVP represses *FT* transcription needed for flowering at lower temperatures (Lee et al., 2007) (Figure 3). However, the levels of *FT* transcripts are higher at higher temperatures, suggesting that SVP is temperature-dependent (Lee et al., 2007).

Furthermore, ambient temperature induction is rooted in known genetic pathways of floral transition, and appears to involve RNA processing (Balasubramanian et al. 2006). *FCA* expression is regulated by ambient temperature both at the transcriptional and post-translational level, and FCA and SVP may be involved in thermosensory regulation of the small RNA miR172 whose targets repress photoperiod-dependent flowering (Jung et al., 2012; Schwab et al., 2012).

The Gibberellic Acid (GA) Pathway

In *Arabidopsis thaliana,* Gibberellins (GAs) promote growth of plant organs and induce phase transitions during development. GAs initiate flowering through the activation of genes encoding floral integrator proteins SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), LEAFY (LFY), and FLOWERING LOCUS T (FT) (Figure 4). GIBERELLIN 20 OXIDASE (GA20ox) enzyme catalyzes the biosynthesis of GA (Mutasa-Gottgens and Hedden, 2009). The bioactive GA (GA_4) accumulates from other parts of the plant, mainly the petioles, at the meristem before floral induction. GA2-oxidase (GA2ox) deactivates GA by hydroxylation (Mutasa-Gottgens and Hedden, 2009). GA signaling pathway is tightly regulated by developmental and environmental cues (Mutasa-Gottgens and Hedden, 2009). Two other major phytohormones abscisic acid and brassinosteroids display only partial interaction with gibberellic acid, which assumes the dominant role in the hormone regulation of the transition from vegetative to reproductive phase of development (Domagalska et al., 2010).

The Circadian Clocks

The photoperiodic induction of flowering involves the interaction between the internal rhythms under control of the plant's circadian clock and the external day-length signals (Bunning, 1960; Pittendregh and Minis, 1964). The circadian clock oscillator genes control oscillation of mRNA accumulation of many transcription factor families, including MYB, bHLH, and bZIP, and affect between 30 - 40% of *Arabidopsis* genes (Alabadi et al., 2001; Davis et al., 2001; Hicks, K.A. et al. 2001; Doyle et al., 2002; Covington et al., 2008; Hanano et al., 2008). The internal

and external times remain synchronized through the process of entrainment, which is a result of the perception of day length by photoreceptors (Jones, 2009).

Circadian clocks are broken down into three component types: input pathways to allow entrainment to environmental conditions, a central oscillator consisting of a negative feedback loop, and output pathways which act to modulate responses dependent on the endogenous cues of the oscillator (reviewed by Jones, 2009) (Figure 5). The core autoregulatory positive/negative feedback loop connects the clock oscillator genes: *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) to *TIMING OF CAB EXPRESSION1* (*TOC1*) (Alabadi et al., 2001). *LHY* and *CCA1* induce MYB-like transcription factors that act additively to repress *TOC1* during the day (Alabadi et al., 2002). TOC1, in turn, activates transcription of *CCA1* (Alabadi et al., 2001). Without this core loop, the circadian clock ceases to function (Ding et al., 2007). The three-loop model incorporates and expands upon the core loop with two additional loops: the morning loop and the evening loop (Locke et al., 2006; Zeitlinger et al., 2006). The morning loop contains PSEUDO-RESPONSE REGULATORs (PRRs) including PRR5, PRR7, and PRR9, repressing *CCA1* and *LHY* expression (Nakamichi et al., 2005, 2010). PRR5 also stabilizes TOC1 protein in nucleus (Wang et al., 2010). CCA1 and LHY reciprocally promote *PRR9* and *PRR7* (Nakamichi et al., 2010). The evening loop involves the ubiquination of TOC1 and PRR5 by ZEITLUPE (ZTL), an F-box protein, together with FKF1 and LKP2 (Mas et al., 2003; Baudry et al., 2010). ZTL then interacts with GI in blue-light to stabilize and negatively regulate active levels of nuclear TOC1 and PRR5 protein (Kim et al., 2007; Kiba et al., 2007).

Three genes have been reported as essential to sustain the circadian clocks in *Arabidopsis thaliana*: *EARLY FLOWERING3* (*ELF3*), *EARLY FLOWERING4* (*ELF4*), and *LUX ARRYTHMO*/*PHYTOCLOCK1* (*LUX*/*PCL1*) (Hicks et al., 1996; Doyle et al., 2002; Hazen et al., 2005; Onai and Ishiura, 2005; McWatters et al., 2007; Thines and Harmon, 2010). LUX, ELF3, and ELF4 are proposed to repress *PRR9* (reviewed by Nakamichi, 2011). Together with ELF4 that represses *PRR7* and *GI*, and with CCA1 that activates *PRR9,* LUX, ELF3, and ELF4 sustain rhythmic expression of *PRR9* (Helfer et al., 2011; Kolmos and Davis, 2007; Kolmos et al., 2009; Dixon et al., 2011; Kolmos et al., 2011; Herrero et al., 2012). They also appear to act in the morning loop to promote levels of *CCA1* and *LHY*, while reducing levels of *TOC1* (Schaffer et al., 1998; Fowler et al., 1999; Davis, 2002; Onai and Ishiura, 2005; McWatters et al., 2007; Thines and Harmon, 2010).

The Photoperiod Pathway

Central genes involved in the downstream of the photoperiod pathway in *Arabidopsis thaliana* are *GIGANTEA*(*GI*), *CONSTANS*(*CO*), and *FLOWERING LOCUS T* (*FT*) (Redei, 1962; Koorneef *et al*., 1991, 1998) (Figure 6). These genes are well conserved across many species of flowering plants (Hecht *et al*., 2005).

CO encodes a putative zinc finger transcription factor, and plays a central role in photoperiodic flowering control in Arabidopsis (Putterill *et al.*, 1995). *CO* is expressed predominantly in the phloem (An *et al*., 2004). The quantity of *CO* expression has a rate-limiting effect upon flowering time, with greater CO accumulation resulting in earlier flowering (Samach *et al*., 2000). Overexpression of *CO* results in an early-flowering phenotype and increased expression of *FT*, a floral

integrator controlling flowering. However, CO induces flowering independent of FT as well (An *et al*., 2004).

CO transcription is under the control of the circadian clock. In *Arabidopsis thaliana, CO* expression is regulated by *GI*, *FLAVIN-BINDING, KELCH REPEAT, F-BOX 1* (*FKF1*), and *CYCLING DOF FACTOR* (*CDF*) genes, which are also regulated by the circadian clock (Fornara et al., 2009; Imaizumi et al., 2005; Sawa et al., 2007) (Figure 6). In long day, FKF1 and GI accumulate as their expression is in phase, but under non-inductive short day GI and FKF1 peak out of phase from one another (Sawa et al., 2007). When in photo-inductive long day, the LOV domain of FKF1 protein binds to the N-terminus of GI in the presence of blue light, which is sufficiently abundant only in long day (Sawa et al., 2007). However, the FKF1-GI complex requires CDF1 in order to regulate *CO* expression (Fornara et al., 2009; Imaizumi et al., 2005). Three homologs of CDF1 play a redundant role in repressing *CO* transcript levels by directly binding to the *CO* promoter (Fornara et al., 2009).

Analysis of the abundance of the three proteins showed that CDF1 and its homologs peak first, followed by GI, and then finally FKF1 peaks in the afternoon under long day conditions (Imaizumi et al., 2005; Sawa et al., 2007). In the morning, CDF1 and its homologs bind to the *CO* promoter (Fornara et al., 2009). Next, GI accumulates and forms a CDF-GI complex to repress *CO* (Fornara et al., 2009). Once FKF1 protein is abundant, it interacts with the CDF-GI complex and targets CDFs for degradation (Imaizumi et al., 2005; Sawa et al., 2007), resulting in the transcription of *CO* (Fornara et al., 2009; Imaizumi et al., 2005; Sawa et al., 2007). Furthermore, plants that carry mutations in *CDF1* and *CDF1* homologs suppressed the late

flowering phenotype caused by *gi* mutation, confirming the redundant function of the CDF1 homologs (Fornara et al., 2009).

CO is also regulated post-translationally. CONSTITUTIVELY PHOTOMORPHOGENIC (COP1) and members of the SUPPRESSOR OF PHYA-105 (SPA) protein family regulate CO protein stability and accumulation (Jang et al., 2008; Liu et al., 2008; Hoecker and Quail, 2001; Laubinger et al., 2006; Laubinger and Hoecker, 2003). CRYPTOCHROME1 (CRY1) is upstream of COP1 and acts as a repressor (Liu et al., 2008). COP1 interacts with CO to act as an upstream negative regulator of CO function by targeting CO for ubiquination (Jang et al., 2008). Similarly, SPA proteins, phytochrome A (phyA) - and other phytochromes - -specific signaling intermediates, elicit degradation of CO by COP1-mediated ubiquination (Hoecker and Quail, 2001; Laubinger et al., 2006; Laubinger and Hoecker, 2003). Furthermore, COP1 along with EARLY FLOWERING 3 (ELF3) affects transcription of *CO* by limiting the accumulation of FKF1-GI in darkness (Yu et al., 2008).

The Photoperiod Pathway and the Phytoreceptors

The circadian clocks are synchronized by cyclical changes in the environment, such as light/dark transitions (Figure 6). Light is among the most important stimuli. The phytoreceptors, from the phytochrome and cryptochrome families, transduce the light signal to the central oscillator to entrain the plant's internal clock (Somers et al., 1998). Genes playing a major role in light perception and light-regulated rhythmic responses are *PHYTOCHROME A* (*PHYA*), *PHYTOCHROME B* (*PHYB*), *CRYPTOCHROME 1* (*CRY1*), and *CRYPTOCHROME 2* (*CRY2*), and all are key genes in the photoperiod pathway of flowering (Somers et

al., 1998). There are five photoconvertible phytochromes (PHYA - PHYE). PHYC, PHYD, and PHYE are mostly redundant in function or act in concert with PHYA and PHYB in development and, particularly, circadian clock entrainment and flowering (Franklin et al., 2003; Clack et al., 2009; Somers et al., 1998, Franklin et al., 2003a). Red light receptors PHYA and PHYB in *Arabidopsis thaliana* act additively to detect high- and low-intensity far-red and red light (Parks, B.M. et al. 1993; Reed, J.W. et al. 1993; Somers et al., 1998; Briggs, W.R. et al. 2001). Cryptochrome proteins, CRY1 and CRY2, act redundantly to perceive blue light, and bind to COP1 to repress its function and up-regulate CO (Ahmad, M. et al. 1993; Lin, C. et al. 1998; Liu et al., 2008; Valverde et al., 2004). Together, these receptors are able to transmit light stimuli into a signal to control flowering (Devlin et al., 2000; Franklin et al., 2003; Franklin et al., 2004, Yanovsky et al., 2002).

Under red light induction, PHYB reduces CO abundance (Putterill, J. et al. 1995; Mockler et al., 1999). Under blue light induction, CRY2 inhibits PHYB and induces flowering (Mockler et al., 1999, Mas et al., 2000). CRY1 acts as a co-promoter with CRY2 of *CO* and *GI* genes (Mockler, T.H. et al. 1999). PHYB and CRY1 interact with each other and mediate crosstalk between the red/far-red and blue/UV photoreceptor pathways (Hughes et al., 2012). Additionally, the PHYTOCHROME INTERACTING FACTOR (PIF) family has been shown to control light dependent gene expression directly and indirectly by binding to photoactivated phytochromes with different affinities and to transduce light signals with varied efficiency (Castillon et al., 2007). Consequently, these photoreceptors

are under the control of the circadian clock by way of a regulatory loop that modulates gating and resetting of the circadian clock (Toth et al., 2001).

Floral Integrators and Floral Meristem Identity

All of the flowering pathways converge to floral integrators including *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*), and induce the expression of floral meristem identity (FMI) genes, including *APETALA 1* (*AP1*), *APETALA 3* (*AP3*), and *AGAMOUS* (*AG*), that control floral organ development (Blazquez, M.A. et al. 2000; Lee, H. et al. 2000). *LEAFY* (*LFY*) gene also plays a critical role in the convergence of all these flowering pathways (Borner, R. et al. 2000). However, *LFY* largely corresponds to flower development, rather than the initiation of flowering (Bernier, G. 1988).

FT is known as a key flowering induction signal called florigen. *FT* expression is gradually induced with maximum expression at floral initiation (reviewed by Turck et al., 2008). In Arabidopsis, *FT* is induced by the *CO* gene in LD (Samach, A. et al. 2000). *FT* also acts downstream of the autonomous and vernalization pathway, and, therefore, is a key flower integrator. Similarly, the expression of *SOC1* is regulated by the light pathway, but also by the autonomous, temperature, and gibberellin pathways (Borner, R. et al. 2000; Hepworth, S.R. et al. 2002). *LFY* functions in part downstream of *SOC1* (Lee, H. et al. 2000).

The Age Pathway

The final pathway, the age pathway, can induce flowering independently of FT function (Figure 7). Wang et al., (2009) showed the accumulation of the

SQUAMOSA PROMOTER BINDING LIKE (*SPL*) MADS box genes increase as the plant ages. SPLs act downstream of FT as transcription factors to induce flowering by initiating expression of *FRUITFULL* (*FUL*), *LFY*, and *SOC1*. SPL proteins are negatively regulated by miRNA-156. The concentration of miRNA-156 is high in young plants and declines as the plants age independent of day length (Wang et al., 2009).

1.3 Photoperiodic Flowering in Short Day Model *Oryza sativa*

Rice is a model system for short-day plants and provides an evolutionarily comparative view of the photoperiodic flowering pathway (Figure 8). The GI-CO-FT pathway, a central module of photoperiodic flowering gene networks in Arabidopsis, is evolutionarily conserved in rice (reviewed in Tsuji et al., 2010). *OsGI*, the rice *GI* homolog, is a circadian regulated gene and an activator of *Hd1* expression (Hayama et al., 2002). Hd1, the rice CO homolog, promotes *Hd3a* expression under SD condition (Hayama et al., 2003). Hd3a, a rice FT homolog, has been shown to act as a florigen in rice, as seen in Arabidopsis (Tamaki et al., 2007).

However, there are differences between Arabidopsis and rice mechanisms and pathways of photoperiod control. First, in non-inductive conditions, the regulatory mechanism of flowering in SD rice deviates from the LD Arabidopsis model (Yano et al., 2000). In LD, Hd1 function converts into a repressor due to the modification of the protein by a phytochrome that reduces *Hd3a* (Ishikawa et al., 2011). Second, Hd1 is not the lone regulator of *Hd3a*. Ehd1, an evolutionary unique

gene that does not have an ortholog in Arabidopsis genome, promotes *Hd3a* through a gene network distinct from the photoperiod gene network in Arabidopsis (Itoh et al., 2010). In addition to *Hd3a*, *RICE FLOWERING LOCUS T1* (*RFT1*) also acts as a florigen under non-inductive LD condition (Komiya et al., 2009). When both Hd3a and RFT1 are suppressed, flowering does not occur, suggesting that flowering is fully dependent on these factors in rice (Komiya et al., 2008). Therefore, the key genes in the photoperiod pathway: *FT*, *CO*, and *GI*, are conserved in *Arabidopsis thaliana* and rice, also in wheat, barley, pea, tomato, onion, and other species (reviewed by Turck et al., 2008). However, the mechanism of their actions may vary as demonstrated by the rice CO homolog, Hd1. Thus, rice provides a unique system to study flowering control.

1.4 Soybean Flowering

Soybean, one of the important oil and protein crops in the world, supplies the major plant protein resource for food, feed, and industrial uses. In soybean, genetic control of flowering time has been used in classical breeding programs for many years and is essential for effective crop production of a given cultivar in specific latitudinal and climatic regions.

In soybean, nine maturity loci, eight E loci (E1 to E8) and J, have been identified and characterized at phenotypic and genetic levels (Bernard 1971; Buzzell 1971; Buzzell and Voldeng 1980; McBlain and Bernard 1987; Bonato and Vello 1999; Cober and Voldeng 2001; Cober et al. 2010). Of these, the E1, E3, E4 and

E7 loci have been found to be related to photoperiod sensitivity under different light quality conditions. The E3 locus causes flowering delays under long days with different light qualities, indicating E3 is less sensitive to light quality (Cober et al. 1996). Whereas, flowering delays caused by the presence of the E1, E4, and E7 alleles under long day require light quality with lower red to far-red (R:FR) quantum ratios (Cober et al. 1996; Cober and Voldeng 2001). Phytochrome regulation of the effect of photoperiod on the development of soybean is in effect even at the post-flowering stage; phytochromes are, therefore, believed to regulate soybean lifetime from emergence to maturity (Han et al. 2006). Among the known E loci, E1 has the largest effect on flowering time under field conditions (Abe et al. 2003; Bernard 1971; Stewart et al. 2003). The E1 gene functions as a putative flowering repressor (Thakare et al. 2010, 2011). Of the genes identified in soybean relating to flowering, E3 and E4 (encoding PHYA), and GmCRY1 are photoreceptors with the function of sensing different light. The three-member families of the circadian clocks, ZTL, FKF1, and LOV-KELCH PROTEIN2 (LKP2), are represented in soybean (Hecht et al. 2005; Quecini et al. 2007). A soybean MADS box gene *Glycine max AGAMOUS Like 2* (*GmGAL2*), a homolog of *AGL11*/*STK*, promoted expression of key flowering genes *CO* and *FT* when ectopically expressed in Arabidopsis (Xu et al., 2010). Soybean *AGL20*/*SOC1*, *Glycine max AGAMOUS Like 1* (*GmGAL1*), was cloned from soybean and its function was investigated in transgenic Arabidopsis lines, and showed similar, but not identical, functions on plant development compared with its homolog in Arabidopsis (Zhong et al., 2012).

Among the soybean CRYPTOCHROME 1 (CRY1) homologs, only GmCRY1a (Glyma 04g11010) has a strong promoting effect in floral initiation; it also oscillates with a circadian rhythm in different photoperiods (Zhang et al. 2008). Association of the expression of GmCRY1a with a latitudinal cline in photoperiodic flowering of soybean has additionally been identified (Zhang et al. 2008). However, there is no flowering time QTL linked to these genes (Matsumura et al. 2009) leaving their function in flowering time to be clarified.

GmPHYA3 and GmPHYA2 have been identified as causal genes of two major flowering loci, E3 and E4 (Figure 8). GmPHYA3 and GmPHYA2 respond differently to different red to far-red quantum (R:FR) ratios, suggesting that they mediate independent but overlapping pathways (Abe et al. 2003; Cerdán and Chory 2003; Cober et al. 1996). Ectopic expression in Arabidopsis has shown that the soybean *PHYB* homolog, *GmPHYB1*, was well conserved both at the level of sequence and physiological function (Wu et al., 2011). Recently, the E2 locus in soybean was identified to encode GmGI, a homolog of Arabidopsis GIGANTEA, and have multiple functions involved in the circadian clock and flowering (Watanabe et al, 2011). Two homologs of FLOWERING LOCUS T (GmFT2a and GmFT5a) were found to coordinately promote photoperiodic flowering in soybean (Kong et al., 2010; Sun et al., 2011). In addition, Dt1 locus was found to encode a homolog of TERMINAL FLOWER1 (TFL1), a well-known FT homolog and a flowering repressor in Arabidopsis (Bradley et al., 1997; Liu et al., 2010).

The soybean genome contains four myb transcription factors: LHY1/CCA1-like genes (Liu et al. 2009). These genes are known to act as the core oscillator of the

circadian clocks in Arabidopsis (see above). Soybean LHY1/CCA1-like genes are expressed with the diurnal rhythms, indicating that they might also function as the core oscillator in soybean (Liu et al. 2009; Thakare et al. 2010, 2011). However, their relationship with flowering time or photoperiod response needs be explored further.

Successful identification of crucial soybean flowering genes and further studies of their regulatory mechanism will be important steps toward deep understanding of photoperiodic flowering responses in soybean.

CHAPTER 2

CHARACTERIZE CO HOMOLOG EXPRESSION IN SOYBEAN BY QRT-PCR

2.1 Background

Flowering is a key trait to angiosperms' survival and reproductive success that also determines the life cycle of the plant. However, there is much to be known in the molecular basis of flowering for important crop species.

In my research, I focused on soybean (*Glycine max* [L.] Merr.). *G. max* is a short-day flowering dicot native to East Asia, and contains the most combined seed protein and oil content of any major crop species. This makes soybean an important staple of human diet. Moreover, legumes including soybean have the capacity to fix atmospheric nitrogen due to a symbiotic relationship with rhizobia (*Bradyrhizobium*). While environmental concerns arise from the use of nitrogen fertilizers, soybean possesses the characteristic benefit of requiring less nitrogen to grow, in addition to its protein-rich nutrition profile. Furthermore, nitrogen application decreased and corn yield was increased by approximately 45 bu./acre if a soybean crop preceded the corn (Rehm et al., 2006).

Like many important crop species, little is known of flowering control in soybean. Eight maturity loci (*E*-loci), E1-E8, have been identified and determined to be involved in varying aspects of soybean flowering and seed maturation (Xia et al., 2012). The *E3* and *E4* genes encode *PHYA* homologs and the *E2* gene encodes a *GI*

homolog (Liu et al., 2008; Shin and Lee, 2012; Watanabe et al., 2011). The aim of my research is to better understand the mechanisms of photoperiodic flowering control in soybean in order to improve soybean adaptability and yield. This work will provide new information on flowering genes and their function that allow modification of the plant's life cycle and greater diversification of cultivation strategies.

The photoperiod pathway involves two key environmental inputs: day length and light quality. Homologs of major components of the photoperiod pathway: GI, CO, and FT (Corbesier et al. 2007); exist in soybean genome (Hecht et al. 2005; Liu et al. 2009; Quecini et al. 2007; Thakare et al. 2010, 2011). However, whether the CO-FT module is conserved in soybean and whether CO plays a crucial function in the flowering are unclear. CO is highly conserved across flowering plants, however, in some plant species, it is known that CO homologs do not have a central function in the control of flowering time (Ben-Naim et al. 2006; Hayama et al. 2007). A large number of CO homologs are present in the soybean genome due to genome duplications (Schmutz et al., 2010). Among these homologs, Thakare et al. (2010) found that Glyma08g28370 and Glyma18g51320 have the highest sequence similarity to the Arabidopsis COL2 (At3g02380) protein (Suarez-Lopez et al. 2001; Hayama and Coupland 2003). The expression of Glyma08g28370 or Glyma13g01290 seems to be arrhythmic or at a very low level (Thakare et al. 2010). Expression of GmCOL4 (Glyma20g07050) and several CO like sequences also has a diurnal rhythm (Huang et al. 2011; Jiang et al. 2011; Liu et al. 2011; Zhang et al. 2010). Recently, *GmCOL9* (*Glyma19g05170*), *GmCOL10* (*Glyma13g01290*), and

GmCOL11 (*Glyma05g35150*) were identified, cloned, and used in an expression analysis. *GmCOL9*, most closely related to *Arabidopsis COL13*, was not regulated by a circadian mechanism, but expression was found throughout the plant, indicating that it may be a multifunctional gene in soybean development (Huang et al., 2011). *GmCOL10*, most closely related to *Arabidopsis COL5* (*At5g57660*), was regulated in a circadian manner under short day condition and expressed in vegetative organs (Liu et al., 2011). *GmCOL11*, most closely related to *A. thaliana COL6* (*At1g68520*) and *COL16* (*At1g25440*), was regulated in a diurnal rather than a circadian manner, and expression was enhanced following flowering, suggesting that it is not directly involved in flowering (Jiang et al., 2011). Overall, the functions of these CO homologs have not been well characterized; in particular, their involvement in flowering time is unknown because there are no QTL linked to GmCOLs.

Short day monocot *Oryza sativa*, rice, contains homologs of long day dicot *Arabidopsis thaliana FT*, *CO*, and *GI*, as well as photoreceptors and circadian clock gene homologs (reviewed in Tsuji et al., 2010). The *FT* homolog in rice, *Hd3a*, is functionally similar to Arabidopsis *FT* that is a floral inducer (Tamaki et al., 2007). The *GI* homolog, *OsGI*, is also functionally similar to Arabidopsis *GI* that up-regulates the *CO* homolog, *Hd1* (Hayama et al., 2002). *Hd1* is also functionally similar to Arabidopsis CO that induces *Hd3a* in flowering-inductive short day (Hayama et al., 2003). However, *Hd1* in rice also functions as a suppressor of *Hd3a* in flowering-non inductive long day (Yano et al., 2000). Therefore, the main flowering genes *FT*, *CO*, and *GI* are conserved in that they are involved in floral transition, however, the function varies as demonstrated by the *CO* homolog, *Hd1*.

In soybean, certain genes of the photoperiod pathway have been identified, such as *PHYA* (*E3/E4*) and *GmGI* (*E2*) (Liu et al., 2008; Shin and Lee, 2012; Watanabe et al., 2011). Recently, *FT* homologs *GmFT2a* and *GmFT5a* have been identified and characterized as putative floral promoters (Sun et al., 2011; Kong et al., 2010). In addition, *E1*, the maturity locus showing the largest effect on flowering time, was demonstrated to repress flowering and delay maturity in soybean (Xia et al., 2012a). However, the function of *CO* homologs has yet to be functionally characterized (Jiang et al., 2011; Huang et al., 2011; Liu et al., 2011). We hypothesize that a *CO* homolog(s) play an important role in soybean photoperiodic flowering, as seen in Arabidopsis and rice. In order to test this hypothesis, the objective of my research is to characterize *CO* homolog expression in soybean by qRT-PCR.

2.2 Materials and Methods

Plant Material and Growth Condition

The USDA Soybean Germplasm Collection, curated by Randall Nelson, provided seed for Clark (PI 548533), Williams 82 (PI 518671), the four near isogenic lines (NILs) (PI 547431; PI 547432; PI 547610; PI 591490), and *Glycine soja* (PI 549046) (Table 1).

The seed was grown within the Plant Sciences Laboratory greenhouse complex. The rooms used were 7E2 and 3E1. Both rooms are located in the Turner Hall greenhouse, and fall under the supervision of Todd Statzer. The Plant Care

Facility (PCF) staff provided watering and pest control. No chemical pesticide was applied at anytime during the experiment, but biological pest control was implemented. 7E2 was equipped with supplemental lighting necessary to grow the plants under long-day (LD) conditions. 3E1 was equipped with a mechanical blackout curtain necessary to grow the plants under short-day (SD) conditions.

12cm square plastic pots were used for the plants. The custom mix, Vodkin mix, was used. The Vodkin mix consists of 2:1 – Soybean Mix : Universal Mix. The Soybean Mix is 1:1:1 – Soil : Perlite : Torpedo Sand, and the Universal Mix is 1:1:1 – Soil : Peat : Perlite. After the soil was autoclaved, it was used to fill the 12cm pots. The pots were placed in either room 7E2 or 3E1.

Pots were separated into pairs. Each pair represents a single set. The sets were labeled with white styrene plant labels. After labeling, the seed was planted in the dry, loose soil approximately 1-1.5 inches deep. A single seed from each of the seven lines were planted randomly between the two pots of the set. Three random lines in one pot and four random lines in the other pot to eliminate variability from additional grow space. The soil was then watered down to induce the germination of the seed. Each set of plants contained one of each line: Clark, Williams 82, the four NILs, and *Glycine soja* (Table 1), for seven plants per set.

30 sets were assigned one of four day-length treatments: SD, LD, Shift I (SI), and Shift II (SII). LD and SD light conditions are 16 and 10 hours, respectively. LD day length lasted from 6:45 until 22:45 with supplemental lighting to extend the day past natural dusk. SD day length lasted from 6:45 until 16:45 with the blackout curtain establishing an artificial night prior to natural dusk. Shift I and Shift II were

germinated and grown in LD for 21 days and shifted to SD for 7 or 12 days, respectively.

All seed was planted 8/13/2010. The sampling for the SD occurred 9/6/2010 (24 DAP). The sampling for LD plants occurred 9/8/2010 (26 DAP). SI and SII were shifted to SD on 9/3/2012 (21 DAP). SI and SII were grown in SD for 7 and 12 days, and then sampled on 9/10/2010 (28 DAP) and 9/13/2010 (31 DAP), respectively.

Each sampling occurred every four hours over a 24-hour period for a total of six samplings. SD and LD sampling occurred every four hours starting at 6:30am with the final sampling occurring at 2:30am the following day. SI and SII sampling occurred every four hours starting at 10:30am with the final sampling occurring at 6:30am the following day. A total of four biological replicates were taken for most samplings, which amounts to 4 sets/day length treatment.

Plant Sampling

The method of sampling for each of the six time points depended upon the day length treatment. Any of the six time points sampled during "day" were done by, first, bringing the appropriate number of sets randomly selected (via MS Excel) to the lab. Second, coolers holding Styrofoam cups labeled with appropriate line number (1-7) and replication (A-D) were filled with liquid Nitrogen. At the time of sampling, the plants were cut to include 3-4 leaf trifoliates (above the cotyledon) and placed into the appropriately labeled Styrofoam cup. The *Glycine soja* lines were similarly sampled, but due to its different morphology all of the plant was

taken for sampling. Each time point was sampled in approximately five minutes as to prevent biochemical/physiological variation since the experiment is extremely time sensitive. The plant samples were then stored at -80°C (-112°F) in labeled aluminum foil. The samples were not freeze-dryed.

Due to missing replicates in our SD treatment, the SD grow was repeated. The same steps to grow and sample were used as explained above for the previous SD treatment. The Repeated Short Day (RSD) plants had four biological replicates. The RSD was planted 6/24/2011 and sampled 7/20/2011 (26 DAP).

The samples were then pulled from -80°C and ground resulting in a homogenous mixture of soybean plant tissue. The ground material was separated into multiple labeled 2mL screw-top tubes. The samples were all kept at low temperatures using liquid nitrogen during grinding and aliquoting of stocks. After the ground material was separated, the tubes were organized and placed back in the -80°C. The material was ground to homogenize the tissue. The material was homogenized since the genes focused on in this study may be found in the leaves, stem (phloem), or SAM. The roots or cotyledons were not sampled.

RNA Preparation

The ground tissue in -80°C was used for RNA isolation. In order to isolate the RNA, 200mg of stock tissue was transferred to a 2mL tube containing a metal lyse ball. The transfers were placed into the TissueLyser II (Qiagen®) to further disrupt the tissue. The transfers were pulverized in liquid nitrogen chilled plates twice for one minute at maximum cycles (30/second).

The next series of events is taken from the RNeasy® Mini Handbook (4 ed.) found in the RNeasy® Plant Mini Kit (Qiagen®) with a few minor modifications: After tissue disruption, 900uL of 10uL β-Mercaptoethanol (β-ME)/1mL Buffer RLT mix were added to each transfer tube for lysis of cells and tissues before RNA isolation. The transfers were immediately mixed vigorously using Vortex-Genie 1® (Scientific Industries®) and placed into a 56°C dry bath for 2-3 minutes. After the short incubation period, the transfer tubes were mixed vigorously and centrifuged for 10 minutes to separate out the plant debris.

The lysate was then transferred to a QIAshredder spin column (lilac) placed in a 2mL collection tube, and centrifuged for 2 minutes at full speed. The supernatant of the flow-through was then transferred to a new 2mL tube. 0.5 volume (450uL) of ethanol was added to the lysate, and mixed by pipetting. The sample was then transferred to an RNeasy® spin column (pink) placed in a 2mL tube, and then centrifuged for 15 seconds. The flow-through was discarded.

The optional On-Column DNase Digestion with the RNase-Free DNase Set was performed on all samples. 350uL of Buffer RW1 was added to the RNeasy® spin column, and centrifuged for 15 seconds. The flow-through was discarded. Next, 10uL DNase I stock solution and 70uL Buffer RDD were added directly to the spin column, and allowed to sit for 15 minutes to allow for DNase digestion. Finally, another 350 uL of Buffer RW1 was added to the RNeasy® spin column. The flow-through was discarded.

500uL of Buffer RPE was added to the RNeasy® spin column, and then centrifuged for 15 seconds. Another 500uL of Buffer RPE was added to the RNeasy®

spin column and centrifuged for 2 minutes. The columns were placed in new collection tubes and centrifuged for 1 minute to remove residual ethanol from the Buffer RPE solution. The columns were then placed in 1.5mL appropriately labeled tubes to collect the eluate containing RNA.

Finally, 30uL of 50°C RNase-free water was added directly to the RNeasy® spin column. The columns were allowed to stand for 1 minute to allow the water to diffuse through the column, and then they were centrifuged for 1 minute to elute the RNA. To increase the concentration of RNA in the solution the eluate was again placed directly on the column, allowed to stand for 1 minute, and centrifuged for 1 minute.

1.5uL of RNA for each sample was taken to the W. M. Keck Center's Functional Genomics Lab (located at the University of Illinois in Urbana-Champaign in 340 of the Edward R. Madigan Lab) for measurement of RNA quality and quantity using the NanoDrop™ 1000 Spectrophotometer (Thermo Scientific®). Each 1.5uL RNA sample was measured and analyzed. The 260/280 ratio was used to indicate the quality of RNA. If the ratio was ≥2.0, then it was accepted as "pure" RNA. The sample concentration was measured in ng/uL based on the Beer-Lambert equation (manipulated to use an extinction coefficient):

$$c = \frac{(A \times e)}{b}$$ [Equation 1]

where c = the nucleic acid concentration in ng/uL, A = the absorbance in Absorbance Units (AU), e = the wavelength-dependent extinction coefficient (40 ng-cm/uL in RNA), and b = is the path length in cm. Sufficient RNA quantity is >500ng/uL for this study (Ingle and Crouch, 1988).

Quantitative RT-PCR

After measuring the RNA quality and quantity successfully, the RNA is used to synthesize cDNA using SuperScript™ III First-Strand Synthesis System for RT-PCR kit (Invitrogen™) and the included protocol. First, 5ug of total RNA is mixed with 1uL of 50uM oligo$(dT)_{20}$ and 1uL of 10mM dNTP mix, which is brought to a total of 10uL with DEPC-treated water depending on the amount of RNA solution required for 5ug of total RNA per sample, at 0°C (on ice). The mix is then incubated at 65°C for 5 minutes, then placed back on ice for at least 1 minute. The mix is centrifuged and the cDNA synthesis mix is then prepared by adding 2uL of 10x RT buffer, 4uL of 25 mM $MgCl_2$, 2uL of 0.1DTT, 1uL of RNaseOUT™ (40 U/uL), and 1uL of SuperScript™ III RT (200 U/uL) to each RNA/primer mixture and centrifuged for a total volume of 20uL. The Eppendorf Mastercycler® pro is used to incubate the cDNA synthesis mix for 50 minutes at 50°C. The reactions are terminated at 85°C for 5 minutes, then placed on ice. The reactions are collected by brief centrifuge, and then 1uL of RNase H is added to each tube and the tubes are incubated for 20 minutes at 37°C, then either stored at -80°C or used immediately for polymerase chain reaction (PCR).

In order to test the cDNA quality, RT-PCR was performed using a primer set for a *PBB2*, a known housekeeping gene that encodes a 20S proteasome beta subunit (Thakare *et al.* 2010). The PCR mixture per reaction is prepared with 1uL of 10x PCR Buffer, 1uL of dNTP (2.5mM), 0.5uL of forward primer (10uM), 0.5uL of reverse primer (10uM), 6.4uL autoclaved H_2O, and 0.1uL of Taq polymerase, in addition to 0.5uL of cDNA in a 0.2mL tube, while on ice, for a total volume of 10uL. The PCR mix

is centrifuged and placed in the Eppendorf Mastercycler® pro for automated thermal cycling.

The PCR performed in this case involved an initialization step at 94°C for 2 minutes. The cycle begins after the initialization step. In cycle, the denaturation step is at 94°C for 20 seconds. The annealing step is at 58°C for 45 seconds. The extension step is at 70°C for 2 minutes. The cycle of these three steps were repeated 30 times. The final elongation step was at 70°C for 5 minutes with a 4°C holding step at the end of the PCR reaction. The reaction was then stored at -20°C or immediately used for gel electrophoresis.

PCR products were visualized using gel electrophoresis to verify the presence of useable cDNA for expression analysis. Gel electrophoresis uses an electric field to sort negatively charged DNA bands amplified via PCR. First, a 1% agarose gel must be made by adding 1g of agarose per 100mL of a 1x TAE solution. The solution was then microwaved to melt the agarose (MP = 85°C). A clear solution indicates the agarose has melted. 5uL of ethidium bromide (EtBr) is then added to the solution, which is a DNA intercalator that fluoresces under UV light for visualization of the DNA bands. The solution is then poured into a gel cast with combs that form the wells. Upon cooling the gel hardens, and can be placed into a gel tank filled with 1x TAE solution that includes EtBr. Once the gel is submerged, 1uL bromophenol blue, a color marker, is added to the PCR reactions, and the reactions are centrifuged. At this point, the PCR reactions can be added into the wells in the gel. Along with the PCR products, a Quick-Load® 2-Log DNA Ladder

(0.1-10.0 kb) (New England Biolabs) is added to the gel to indicate the size of the band.

After the run, the gel can be visualized using Ultra-Violet (UV) light due to EtBr. The Gel Doc™ XR+ (BIO-RAD) was used for UV visualization, and images were analyzed using the Image Lab™ software included. The presence of a *PBB2* band indicates and assures the quality of the cDNA. Quality cDNA is necessary for the qRT-PCR expression profiling of a gene.

The qRT-PCR experiments used FastStart Universal SYBR Green Master {ROX} (Roche Applied Science®). The FastStart is a ready to use master mix that includes all reagents except primers and template. FastStart qRT-PCR reactions require 0.5uL of template, 0.3uL of forward primer, 0.3uL of reverse primer, 7.5uL FastStart Master Mix, and 6.4uL of autoclaved water. The cDNA template is added to a 384-well Clear Optical Reaction Plate (Applied Biosystems®), and then the Master Mix (including primers, FastStart, and dH_2O) is added to each well. At this point, the plate was run using the 7900 HT Fast Real-Time PCR System (Applied Biosystems®) located in the Keck Center. 7900 HT RT-PCR System software recorded the data electronically.

In the experiment, the 384-well plates were loaded with a single biological replicate including all four day lengths treatments (SD, LD, SI, SII), six time points, and seven genotypes, which required 168 wells. In the case of our RSD experiment, six time points, seven genotypes, and three biological replicates were loaded into the qRT-PCR plates, which required 126 wells. The size of the plate allowed for two technical replicates in both cases.

To analyze the data from the experiment, a control is required. The control we used was the gene *PBB2* because it is known to provide constitutive levels of expression (Thakare et al., 2010) (Table 2). For all treatments (SD, LD, SI, SII, and RSD), a control plate with *PBB2* primers was run. The data from the control was used to quantify the results for the experimental data using the comparative C_t method. The comparative C_t method compares C_t values from the genes of interest (GOI) data with the control data to normalize the expression data. The comparative C_t method is also known as the $2^{-[delta][delta]Ct}$ method, where

$$[delta][delta]C_t = [delta]C_{t,sample} - [delta]C_{t,reference} \quad \text{[Equation 2]}$$

where $[delta]C_{t,sample}$ = the C_t value for any sample normalized to the control gene and $[delta]C_{t,reference}$ = the C_t value for the normalizing control (ABI, 2001). The $[delta][delta]C_t$ method is only valid if the amplification efficiencies of the genes of interest and the housekeeping gene are similar. Thakare *et al*. (2010) showed that *PBB2* is similar in amplification efficiency to the genes involved in the study. The data analyzed via the comparative C_t method can be used to show expression profiles for genes of interest.

Identification of *CONSTANS* Homologs

The *CONSTANS* (*CO*) homologs used in this study for expression analysis were identified through phylogenetic analysis. First, the *Arabidopsis thaliana* CO, a protein known to be involved in the photoperiod flowering pathway, amino acid sequence was identified on the phytozome.net website. A BLASTp of CO over the *A. thaliana* proteome located on phytozome.net website identified sixteen CO-like

proteins. A BLASTp of the seventeen Arabidopsis amino acid sequences over the *Glycine max* genome identified 9 CO-like proteins (Table 3).

Phylogenetic Analysis

The amino acid sequences for all 28 genes (Table 3 - 17 *A. thaliana;* 9 *G. max*; 1 *Chlamydomonas reinhardtii*; 1 *Oryza sativa*) are aligned using multiple sequence alignment software, MAFFT. The multiple sequence alignment was manually edited to eliminate loops – portions of the protein where some have sequence, and others large gaps – and ensure orthologous positions are aligned (Table 4). The MEGA (Molecular Evolutionary Genetics Analysis) 5.05 software was used to obtain a maximum likelihood phylogenetic tree (Tamura et al., 2011). The evolutionary history was inferred by using maximum likelihood based on the JTT matrix-based model (Jones et al., 1992). MEGA 5.05 bootstrapped the tree to provide quantifiable understanding of the protein relatedness.

The phylogenetic analysis identified 9 *CO* genes to study. In order to examine the 9 genes identified, primer sets were developed and tested (with the help of Valpuri Sovero, a Post-Doctoral researcher in Yoshie Hanzawa's Lab at University of Illinois in Urbana-Champaign) (Table 2). When developing primers it is important take into account many variables, such as melting temperature, G-C content, and propensity to form dimers. The literature reveals that some of the CO-like homologs in this study have been identified, and qRT-PCR primers already exist for them (Thakare et al., 2010; Huang et al., 2011; Liu et al., 2011; Jiang et al., 2011). However, for those homologs not found in the literature, software has been

developed to help in designing a suitable primer. In this study, Primer3, a free online software, was used to design primers (Rozen and Skaletsky, 2000). The primer information was then sent out for synthesis. Upon return, the primers were tested via PCR using cDNA template from one the many soybean samples available.

2.3 Results and Discussion

A BLAST (Basic Local Alignment Search Tool) using the *Arabidopsis thaliana* CONSTANS (AtCO) amino acid sequence over the *A. thaliana*, *Oryza sativa*, and *G. max* genome on phytozome.net produced a number of CO homologs for candidate gene study. The BLAST over the Arabidopsis genome resulted in 16 CO-like proteins: AtCOL1-AtCOL16 (Table 3). The BLAST over the rice genome resulted in the rice CO homolog *Oryza sativa Heading date 1* (Os*Hd1*), a functionally well-characterized protein (Table 3). The BLAST over the soybean genome resulted in a number of hits, and the nine soybean CONSTANS homologs chosen had the highest identity to *A. thaliana* CO: Glyma18g51320 (COL2a), Glyma08g28370 (COL2b), Glyma19g05170 (COL9), Glyma13g07030 (COLX), Glyma13g01290 (COL5), Glyma17g07420 (COLY), Glyma04g06240 (COL4), Glyma06g06300 (COLZ), and Glyma19g39460 (COL92) (Table 3).

Phylogenetic analysis was carried out to clarify sequence similarity and relatedness of the CO homologs from *Arabidopsis thaliana*, *Oryza sativa*, and *Glycine max* relative to Arabidopsis CO. The results of the phylogenetic analysis revealed the *Glycine max* CO homologs were separated into three major clades (Figure 9). Eight of

the *G*. *max* CO homologs appear to constitute four pairs of homeologous proteins (Glyma18g51320 (COL2a) and Glyma08g28370 (COL2b), Glyma19g05170 (COL9) and Glyma13g07030 (COLX), Glyma13g01290 (COL5) and Glyma17g07420 (COLY), Glyma04g06240 (COL4) and Glyma06g06300 (COLZ)), which most likely correspond to the two ancient genome duplication events that have occurred in soybean (Schmutz et al., 2010). Two such pairs were nested in a single clade (Figure 9). This first clade includes AtCO and OsHd1 with the two pairs of *G*. *max* homologs: COL2a and COL2b, COL9 and COLX. The second clade nested in the tree next to the clade containing AtCO included two pairs of *G*. *max* homologs: COL5 and COLY, COL4 and COLZ. The third clade was located next to the second clade and further from the first clade, and it contained the unpaired AtCO homolog COL92. The position of the nine selected *G*. *max* CONSTANS-like homologs infers a closer evolutionary relationship to *A. thaliana* CONSTANS. The inferred relationship is based on the proximity of the *G. max* CONSTANS-like homologs compared to *A. thaliana* CONSTANS-like genes relative to *A. thaliana* CO. AtCOL1 and AtCOL2 were classified into clade 1, and AtCOL3, AtCOL4, and AtCOL5 were classified into clade 2, whereas AtCOL6-AtCOL16 were classified into clade 3 or an even more distant fourth clade, which contained no soybean homologs.

To understand the functions of *G. max* CO homologs, mRNA expression patterns were characterized using qRT-PCR. The expression analysis required the sampling of soybean plants for RNA extraction. The material used in the experiment was provided by the USDA Soybean Germplasm Collection located at the University of Illinois in Urbana-Champaign. Seven lines were selected. Two lines are U.S.

cultivars: Clark and Williams 82. Four of the lines are NILs with Clark as the recurrent parent: L65-3366, L66-432, and L74-441. The seventh line is a wild ancestor of *Glycine max*, *Glycine soja*. The NILs were selected based on their allelic variability at one of the eight maturity loci (E-loci) (Table 1).

In order to simulate long day (LD) and short day (SD) conditions, the lines were grown in either 10 hours of light (SD) or 16 hours of light (LD). The plants were sampled at six different time points over a 24 time period three weeks after planting. The six time points were: 6:30, 10:30, 14:30, 18:30, 22:30, and 2:30; where, dawn occurred at 6:45 in both day length treatments, and dusk occurred at 16:45 and 22:45 in SD and LD, respectively. Three biological replications were sampled for each time point and day length condition (Figure 10).

The qRT-PCR results across genotypes are not evaluated in this chapter. The differences across genotypes are evaluated in the following chapter using the more reliable RNA sequencing results. In this chapter, the qRT-PCR results are used to determine a difference in expression patterns between day lengths.

The subsequent qRT-PCR experiment was conducted using the material from the aforementioned treatments. Prior to qRT-PCR, primers were designed particular to each gene of interest (Table 2). The primers were designed using the transcript sequences for each soybean *CO* homolog (*www.phytozome.net*). Due to the high similarity of sequences, especially between homeologs, primers were mostly designed using the less conserved, and, thus, more unique, 3' end. The housekeeping gene, *20S PROTEASOME BETA-SUBUNIT 2* (*PBB2*), used for relative expression analysis was selected based on its constitutive expression across all treatments

compared to other housekeeping genes tested, along with similar use in the literature (Thakare et al., 2010). The qRT-PCR experiment was performed in the W.M. Keck Center for Comparative and Functional Genomics, Functional Genomics Division, using an ABI platform.

First, two closely related pairs of *G. max CO* homologs in the first clade: *COL2a* and *COL2b*, and *COLX* and *COL9*; were examined (Figure 9). *COL2a* expression in SD (Figure 11A) begins late night (2:30), and peaks before dawn (6:30), then sharply declines over the next four hours (10:30). By the afternoon (14:30), expression is relatively non-existent and remains faint after dusk and into late night (18:30-22:30). *COL2a* expression in LD (Figure 11B) increases before dusk (22:30) and peaks after dusk (2:30), but declines before dawn (6:30) and is relatively non-existent through the day until before dusk (10:30-18:30). The expression pattern suggests that *COL2a* is expressed differently in SD and LD.

COL2b expression in SD (Figure 11C) begins late night (2:30), peaks before dawn (6:30), and then sharply declines over the next four hours (10:30). By the afternoon (14:30), expression is relatively non-existent and remains faint after dusk and into late night (18:30-22:30). The expression of *COL2b* in LD (Figure 11D) increases before dusk (22:30) and peaks after dusk (6:30), but declines before dawn (6:30) and is relatively non-existent through the day until before dusk (10:30-18:30). The SD and LD expression patterns of *COL2b* are identical to that of *COL2a*, the homeolog of *COL2b.* The results suggest that both *COL2a* and *COL2b* respond to day length.

COL9 expression in SD (Figure 12A) peaks after dusk (18:30). The expression levels decline before dawn (22:30-6:30) and remain faint throughout the daytime (10:30-14:30). The expression pattern suggests that the gene is responding to day length. Expression of *COL9* in LD (Figure 12B) is very faint throughout the 24-hour light cycle. *COLX* expression in SD (Figure 12C) peaks after dusk (18:30). The expression levels decline before dawn (22:30-6:30) and remain faint throughout the daytime (10:30-14:30). *COLX* expression in LD (Figure 12D) appears to peak at two points – mid-day (18:30) and at night (2:30). The expression patterns of the homeologs *COL9* and *COLX* in SD are identical and very similar in LD. The results suggest that both *COL9* and *COLX* respond to day length.

Second, two closely related pair of *G. max CO* homologs in the next clade: *COL5* and *COLY, COL4* and *COLZ*; were examined (Figure 9). *COLZ* expression in SD (Figure 13A) peaks after dusk (18:30), and shows faint levels of expression during the rest of the night with relatively no expression during the day (22:30-14:30). *COLZ* expression in LD (Figure 13B) peaks after dusk (2:30) with a steady decline in expression through the night and into the day (6:30-22:30). *COL4* expression in SD (Figure 13C) peaks before dawn (6:30) and declines through the day and into the night (10:30-22:30). By late night (2:30), expression begins to accumulate again. *COL4* expression in LD (Figure 13D) shows much variation between genotypes, but the accumulation of mRNA appears to occur between late night and early morning (2:30-10:30). The homeologs, *COL4* and *COLZ*, do not show similar patterns of expression in SD. The expression patterns of *COL4* across SD and LD suggest that

COL4 does not respond to day length. The expression patterns of *COLZ* across day lengths suggest that *COLZ* responds to day length.

Expression of *COLY* in SD (Figure 14A) appears to have a steady level of expression beginning before dawn and after dusk (6:30-18:30) with slight elevation of transcription occurring during the day (10:30-14:30). *COLY* expression in LD (Figure 14B) peaks before dawn (6:30) and continues into the early morning (10:30). *COL5* expression in SD (Figure 14C) begins to accumulate before dawn (6:30) and peaks during the day (10:30-14:30) before expression tapers off after dusk (18:30). *COL5* expression in LD (Figure 14D) accumulates at low levels after dusk (2:30) and remains level at dawn (6:30) and into early morning (10:30). The homeologs *COLY* and *COL5* have similar expression patterns across SD and LD. The expression patterns suggest that both homeologs respond to day length.

Third, the *G. max* homolog in the most distant clade, *COL92*, was examined (Figure 9). *COL92* expression in SD (Figure 15A) peaks after dusk (18:30) and shows noticeable levels of transcription four hours later (22:30), but expression is relatively non-existent by late night (2:30) and remains faint into and through the day (6:30-14:30). *COL92* expression in LD (Figure 15B) peaks mid-day (14:30-18:30), but relative expression is much lower in LD than photo-inductive SD. A drop-off in expression occurs after dawn (10:30) and before dusk (22:30). Unlike the other eight genes, *COL92* did not have a homeologous pair. However, the expression pattern of *COL92* suggests that the gene responds to day length.

In conclusion, three of the four pairs of homeologs showed similar expression patterns, suggesting that homeologs resulting from genome duplication

retain similar regulatory mechanisms. However, in the case of *COL4* and *COLZ*, homeologs showed a divergence in expression pattern, suggesting possible evolutionary change in regulatory mechanisms. Eight of the nine *CO*-like homologs demonstrated a response to day lengths. This result suggests that these homologs are involved in photoperiod response.

Two sets of homeologs that have higher sequence similarity to Arabidopsis *CO* than other homologs showed contrasting expression patterns under SD: *COL2a* and *COL2b* peak before dawn, and *COL9* and *COLX* peak after dusk. We found that the peak of *COL2a* and *COL2b* overlapped with the peak of *GmFT* homologs (Kong et al., 2010) (Figure 16A&16B)). Based on the overlapped expression and the high similarity to Arabidopsis *CO*, it is likely that *COL2a* and *COL2b* are important regulators of photoperiodic flowering in soybean.

CHAPTER 3

CHARACTERIZE CO HOMOLOG EXPRESSION IN SOYBEAN BY RNA-SEQUENCING

3.1 Background

Eight loci (E1 - E8) have been characterized and the E1, E3, and E4 loci are photoperiod sensitive (Liu et al., 2008; Xia et al., 2012). E3 and E4 genes encode *Phytochrome A* (*PhyA*) homologs (Liu et al., 2008). Recently, the maturity locus E2 was identified as a *GIGANTEA* ortholog (Watanabe et al., 2011). The recessive e2 exhibited elevated expression of a soybean *FT* homolog, indicating that the function of *GI* is conserved in soybean (Watanabe et al., 2011). Our understanding of the actions of E loci in photoperiodic flowering, however, is still limited.

Sanger-based sequencing technologies provided a tool to read and, subsequently, decipher the genetic code of organisms (Sanger et al., 1977). However, the technique is expensive and time-consuming (Wheeler et al., 2008). Presently, the second generation of sequencing technology represents the main method for genetic discovery on a genome-wide scale with single base precision (as reviewed by Mardis, 2008). High throughput platforms are capable of rendering large amounts of short DNA sequence reads (Edwards and Batley, 2010). Currently, third generation sequencing technologies are being testied that offer improvements on the second generation sequencing technology for obtaining a genome of an organism at lower cost and higher efficiency (Harris et al., 2008; Eid et al., 2009). The second generation sequencing technology has been applied to the study of

transcriptomics that allows improved understanding of genetic pathways. The transcriptomics measures transcript abundance of genes in the form of mRNAs and small RNAs. As the transcriptome is a single recorded moment of the organism's current molecular activity, it is capable of capturing how the static genetic code interfaces with the internal and external environment of the organism to elicit a morphological, physiological, or other biological response. Recently, the draft genome of soybean became available, allowing the second-generation sequencing based transcriptome approach in soybean (Schumtz *et al*., 2010).

The use of qRT-PCR provides useful data in the analysis of gene expression. The results presented in Chapter 2 in this thesis provided useful information for better understanding of the possible function of *G. max CO* homologs. However, I have observed that the qRT-PCR results often show a large variation across biological and experimental replicas. This makes accurate analysis of subtle differences in gene expression difficult. For example, while the qRT-PCR results presented in Chapter 2 provided general expression patterns of *G. max CO* homologs, it was not possible to determine the effects of E loci on *G. max CO* homologs due to large experimental errors. To overcome this problem, a RNA sequencing approach was used in my research. RNA sequencing serves two purposes: 1) underpinning the conclusions drawn from the qRT-PCR analysis on *G. max CO* homologs, and 2) providing a more confident look at the effects of E loci on *G. max CO* homologs and large-scale gene expression patterns.

3.2 Materials and Methods

Plant Material and Growth Condition

The USDA Soybean Germplasm Collection provided soybean seed as previously described (**2.3 Materials and Methods**).

RNA Preparation

RNA was prepared from selected samples for RNA-Sequencing. RNA preparation was done as previously described (**2.3 Materials and Methods**). 189 samples representing 3 day length treatments (SD, LD, Shift I); 3 time points (6:30, 14:30, and 22:30); 7 genotypes (Table 1); and, 3 biological replications) The sampling time points were chosen based on the literature that revealed that two *FT*-like homologs in soybean, *GmFT2a* and *GmFT5a*, were characterized as being involved in SD photo-induction of flowering (Kong *et al*. 2010).

Submission to the Keck Center's Sequencing and Genotyping Lab at the University of Illinois in Urbana-Champaign for RNA-Sequencing required 5ug of total RNA. The volume of total RNA was based on the quantity of RNA (ng/uL) in each sample. Previous to submission, the RNA was tested for degradation using electrophoresis on a 1% Agarose gel. The 189 samples were submitted in 1.5uL tubes containing 40uL of dH_2O, along with the RNA, as instructed by the director of the Sequencing and Genotyping Lab, Alvaro Hernandez.

RNA Sequencing Data Analysis

The samples selected for RNA sequencing were submitted to the W.M. Keck Center for Comparative and Functional Genomics, Sequencing and Genotyping Division. Next, the data rendered by RNA-Seq on the Illumina platform was checked (performed by Waseem Haider, Ph.D. candidate in Yoshie Hanzawa's lab at University of Illinois at Urbana-Champaign) for good quality using two softwares, FastX (Hannon Lab - http://hannonlab.cshl.edu/fastx_toolkit/index.html) and fastQC (Barbaraham Lab http://www.bioinformatics.bbsrc.ac.uk/projects/download.html#fastqc). Next, the individual 100 nucleotide reads were mapped using Bowtie (Langmead et al., 2009) aligned to the soybean transcriptome (Williams 82) obtained from phytozome.net (Schmutz et al., 2010). The cleaned up data was used for analysis (Table 5). For this project, RPKMs of *CO* and *FT* genes were obtained from the cleaned data, and mean and standard error were calculated to make comparisons across time points and genotypes.

3.3 Results and Discussion

To verify the qRT-PCR results and to obtain more accurate gene expression levels, RNA sequencing was carried out. The same plant material (Table 1) grown under the same day length conditions (Figure 10) from the qRT-PCR experiment were used. However, due to the cost constraints of RNA sequencing only three time points were selected: 6:30, 14:30, and 22:30 (Figure 17). The time points were selected for a few reasons. First, selecting time points that were not consecutive made sense in order to get a better idea of expression across the whole day, and not just a segment of the day. Second, the 6:30 time point

ensured that there was a time point in dark (dawn at 6:45 in SD and LD) for both day length treatments. Third, a preliminary analysis of qRT-PCR results and findings in the literature seemed to indicate that these time points were more desirable than the alternative (10:30, 18:30, and 2:30) (Kong et al., 2010; Thakare et al., 2010). However, in the absence of RNA sequencing data to support qRT-PCR data, the analysis is dependent upon the qRT-PCR results only.

The samples selected for RNA sequencing were submitted to the W.M. Keck Center for Comparative and Functional Genomics, Sequencing and Genotyping Division. 189 samples (3 day length treatments; 3 time points; 7 genotypes; and, 3 biological replications) containing 5uL of RNA solution tested for degradation using electrophoresis were submitted. Next, the data rendered by RNA-Seq on the Illumina platform was checked (performed by Waseem Haider, Ph.D. candidate in Yoshie Hanzawa's lab at University of Illinois at Urbana-Champaign) for good quality using two softwares, FastX (Hannon Lab - http://hannonlab.cshl.edu/fastx_toolkit/index.html) and fastQC (Barbaraham Lab http://www.bioinformatics.bbsrc.ac.uk/projects/download.html#fastqc). Next, the individual 100 nucleotide reads were mapped using Bowtie (Langmead et al., 2009) aligned to the soybean transcriptome (Williams 82) obtained from phytozome.net (Schmutz et al., 2010). The cleaned up data was used for analysis (Table 5 & 6). For this project, RPKMs of *CO* and *FT* genes were obtained from the cleaned data, and mean and standard error were calculated to make comparisons across time points and genotypes.

The RPKMs for two closely related pairs of *G. max* homologs in the first clade: *COL2a* and *COL2b*, *COLX* and *COL9*, were obtained from the RNA sequencing data (Figure 4). *COL2a* and *COL2b* RNA-seq results showed expression patterns similar to/consistent with the

qRT-PCR results (Figures 6&12). *COL2a* expression in SD (Figure 18A) shows Clark (blue – Table 1) and Williams 82 (dark green - Table 1) to have a similar number of RPKMs (reads per kilobase of exon model per million mapped reads) (Figure 18A). At the 6:30 time point, the three lines, L65-3366 (yellow – Table 1), L66-432 (purple –Table 1), and L92-1195 (light green – Table 1), display expression levels slightly lower in RPKMs than Clark and Williams 82. At the 6:30 time point, the two lines, L74-441 (orange – Table 1) and *Glycine soja* (red - Table 1), display RPKMs higher than Clark (blue) and Williams 82 (dark green), and an even higher amount of RPKMs than L65-3366 (yellow), L66-432 (purple), and L92-1195 (light green). The results suggest that the NILs and *Glycine soja* genotypes express *COL2a* differently in SD.

COL2a expression in LD (Figure 18B) shows Clark (blue) has a higher number of RPKMs compared to Williams 82 (dark green) (Figure 18B). The RPKMs for L66-432 (purple) and *Glycine soja* (red) are less than the number of reads for Clark and Williams 82. While the RPKMs for L65-3366 (yellow) and L92-1195 (light green) fall between the two wild types, the RPKMs for L74-441 (orange) are greater than Clark. The results suggest that all other genotypes aside from L74-441 (orange) express *COL2a* differently in LD.

COL2b expression in SD (Figure 18C) at the 6:30 time point shows Clark (blue) and Williams 82 (dark green) to not be significantly different from each other and L66-432 (purple). Lines L74-441 (orange), L92-1195 (light green), and *Glycine soja* (red) RPKMs, while not significantly different than Clark, are significantly greater than Williams 82. L65-3366 (yellow) is significantly less than both Clark and Williams 82 at the 6:30 time point. The results suggest that *COL2b* is expressed differently in SD in the L65-3366 (yellow) genetic background.

COL2b expression in LD (Figure 18D) appears to be significantly different at the 22:30 time point, but not at the 6:30 time point, between Williams 82 (dark green) and Clark (blue). L92-1195 (light green) and L74-441 (orange) appear to be the only lines significantly different than the other lines and between each other at the 6:30 time point. The 22:30 time point shows that L65-3366 (yellow), L66-432 (purple), L92-1195 (light green), and *Glycine max* (red) are significantly different from Clark and Williams 82. The results suggest that *COL2b* is expressed differently in the NIL backgrounds than the recurrent parent Clark in LD.

COLX and *COL9* showed expression patterns similar to/consistent with the qRT-PCR results (Figures 12 & 19). *COL9* expression in SD (Figure 19A) at 6:30 time point shows Clark (blue) and Williams 82 (dark green) are significantly different. The L74-441 (orange), L92-1195 (light green), and *Glycine max* (red) lines appear to be significantly different from the recurrent parent Clark, but the RPKMs are extremely low at the 6:30 time point. The 22:30 time point shows slightly higher and more even expression across all genotypes with no significant difference between lines. The results suggest that *COL9* expression across genotypes is not significantly different.

COL9 expression in LD (Figure 19B) at all time points is not significantly different between Clark (blue) and Williams 82 (dark green). The L92-1195 (light green) and *Glycine max* (red) lines show expression significantly different from Clark. At the 14:30 and 22:30 time points, all NILs and *Glycine max* genotypes are significantly different than Clark. The results suggest that *COL9* in LD is expressed differently across genotypes. However, RPKMs of *COL9* in LD and SD are very low, so the difference in expression may not be significant enough to elicit a different response in the plant.

COLX expression in SD (Figure 19C) at the 6:30, but not 10:30 or 22:30, time point shows significant difference between the mapped reads between Clark (blue) and Williams 82 (dark green). Lines L66-432 (purple), L74-441 (orange), and L92-1195 (light green) are significantly different at the 6:30 time point, but not the 22:30 time point. The results suggest that *COLX* in SD is expressed differently across genotypes.

COLX expression in LD (Figure 19D) at the 6:30 and 14:30, but not 22:30, time points between Clark (blue) and Williams 82 (dark green) are significantly different. At 6:30, the L92-1195 (light green) and *Glycine max* (red) expression are significantly different than Clark. At the 14:30 time point Clark is significantly different than all other genotypes. At 22:30, L66-432 (purple), L74-441 (orange), L92-1195 (light green), and *Glycine soja* (red) are significantly different than the NIL recurrent Clark parent. Results suggest that *COLX* in LD is expressed differently across genotypes. However, RPKMs of *COL9* and *COLX* were considerably lower, especially relative to *COL2a* and *COL2b*, so this fact may suggest that the differences are not enough to elicit a variable response in the plant.

The number of mapped reads for the *G. max FT* homologs, *GmFT2a* and *GmFT5a*, were obtained from the RNA sequencing data to compare the expression variation across genotypes (Table 6). *GmFT2a* expression in SD (Figure 20A) at 6:30, but not 14:30 or 22:30, time point is not significantly different between Clark (blue) and Williams 82 (dark green). Clark is significantly different at the 6:30 and 14:30 time point than all other genotypes. At time point 14:30, there is not much variation between Williams82, L65-3366, L66-432, L74-441, and L92-1195; however, Clark shows ~30% more reads than these five lines. The results suggest that *GmFT2a* in SD is expressed differently across genotypes.

GmFT5a expression in SD (Figure 20B) at all time points show RPKMs for Clark to be significantly different than Williams 82 (dark green). At the 6:30 time point, Clark has significantly higher RPKMs than all other genotypes. At time point 6:30, there is not much variability between Williams82, L65-3366, L66-432, L74-441, and L92-1195; however, Clark shows 20-30% more reads than these five lines. At the same time point, Clark shows ~50% more reads than *Glycine soja* (red). Otherwise, the expression levels appear consistent at the other time points. The results suggest that *GmFT5a* in SD is expressed differently across genotypes.

In conclusion, RNA-sequencing showed smaller variation between biological replications than what was observed in the experimental replications observed using qRT-PCR. Therefore, RNA-sequencing may be a more appropriate approach to investigate the subtle genetic effects at the level of transcription.

COL2a and *COL2b* did not show obvious differences across genotypes, suggesting the E-loci: E1, E2, E3, and E5, and genetic background of *G. soja* may not affect *COL2a* and *COL2b* expression. Two possible scenarios are: 1) E1, E2, E3, and E5, and genetic background of *G. soja* may regulate flowering in a CO-independent manner, 2) E1, E2, E3, and E5, and genetic background of *G. soja* may control CO through post-transcriptional regulation. The latter scenario fits well with a model of CO regulation known in Arabidopsis, where Phytochromes and Cryptochromes regulate CO protein stability (Fornara et al., 2009; Imaizumi et al., 2005; Sawa et al., 2007). Supporting this scenario, *GmFT*s, potential downstream genes of *COL2a* and *COL2b*, show variation in mRNA expression across genotypes. Further analysis on COL2a and COL2b protein regulation will clarify this question.

Literature Cited

Abe, J., Xu, D.H., Miyano, A., Komatsu, K., Kanazawa, A., Shimamoto, Y. (2003) Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. Crop Sci 43:1300–1304

Alabadi, D., Oyama, T., Yanovsky, M., Harmon, F., Mas, P., Kay, S. (2001) Reciprocal regulation between *TOC1* and *LHY*/*CCA1* within the *Arabidopsis* circadian clock. Science 293:880-883

Alabadi, D., Yanovsky, M., Mas, P., Harmer, S., Kay, S. (2002) Critical role for CCA1 and LHY in maintaining circadian rhythmicity in *Arabidopsis*. Current Biology 12:757-761

An, H., Roussot, C., Suarez-Lopez, P., Corbesier, L., Vincent, C., Pineiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C., Coupland, G. CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. Development 131:3615-3626

Balasubramanian, S., Sureshkumar, S., Lempe, J., Weigel, D. (2006) Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. PLoS Genetics 2:e106

Baudry, A., Ito, S., Song, Y., Strait, A., Kiba, T., Lu, S., Henriques, R., Pruneda-Paz, J., Chua, N., Tobin, E., Kay, S., Imaizumi, T. (2010) F-Box proteins FKF1 and LKP2 act in concert with ZEITLUPE to control *Arabidopsis* clock progression. The Plant Cell 22:606-622

Bäurle I, Dean C (2006) The timing of developmental transitions in plants. Cell 125: 655–664

Bäurle I, Dean C (2008) Differential Interactions of the Autonomous Pathway RRM Proteins and Chromatin Regulators in the Silencing of Arabidopsis Targets. PLoS ONE 3(7): e2733. doi:10.1371/journal.pone.0002733

Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., Samach, A., Lifschitz, E. (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. The Plant Journal 46:462-476

Bernard, R.L. (1971) Two major genes for time of flowering and maturity in soybeans. Crop Science 11:242-244

Bernier, G. (1988) The control of floral evocation and morphogenesis. Ann. Rev. Plant Physiol. Plant Mol. Biol. 39:175-219

Blazquez, M. (2000) Flower development pathways. Jour. Cell Sci. 113:3547-3548

Blazquez, M., Ahn, J., Weigel, D. (2003) A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. Nature Genetics 33:168-171

Bonato, E.R., Vello, N.A. (1999) *E6*, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 22:229–232

Borner, R. et al. (2000) A MADS domain gene involved in the transition to flowering in Arabidopsis. Plant J. 24:591–599

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., Coen, E. (1997) Inflorescence commitment and architecture in *Arabidopsis*. Science 275:80-83

Briggs, W.R., Beck, C.F., Cashmore, A.R., Christie, J.M., Hughes, J., Jarillo, J.A., Kagawa, T., Kanegae, H., Liscum, E., Nagatani, A., Okada, K., Salomon, M., Rudiger, W., Sakai, T., Takano, M., Wada, M., Watson, J.C. (2001) The phototropin family photoreceptors. Plant Cell 13:993-997

Bünning, E. (1960). Circadian rhythms and the time measurement in photoperiodism. Cold Spring Harb. Symp. Quant. Biol. 25, 249–256

Buzzell, R.I. (1971) Inheritance of a soybean flowering response to fluorescent-daylength conditions. Can. Jour. Genet. Cytol. 13:703–707

Buzzell, R.I., Voldeng, H.D. (1980) Inheritance of insensitivity to long daylength. Soybean Genet. Newsletter 7:26–29

Castillon, A., Shen, H., Huq, E. (2007) Phytochrome interacting factors: central players in phytochrome-mediated light signaling networks. Trends in Plant Science 12:514-521

Cérdan, P., Chory, J. Regulation of flowering time by light quality. Nature 423:881-885

Chandler J, Dean C (1994) Factors influencing the vernalization response and flowering time of late flowering mutants of Arabidopsis thaliana (L.) Heynh. J Exp Bot 45:1279–1288

Chouard, P. (1960) Vernalization and its relations to dormancy. Annu. Rev. Plant Physiol. 11:191-238

Clack T, Shokry A, Moffet M, Liu P, Faul M, Sharrock R.A. (2009) Obligate heterodimerization of Arabidopsis phytochromes C and E and interaction with the PIF3 basic helix-loop-helix transcription factor. The Plant Cell 21:786-799

Cober, E., Tanner, J., Voldeng. H. (1996) Soybean photoperiod-sensitivity loci respond differentially to light quality. Crop Science 36:606-610

Cober, E., Voldeng, H. (2001) Low R:FR light quality delays flowering of *E7E7* soybean lines. Crop Science 41:1823-1826

Cober, E., Molnar, S., Charette, M., Voldeng, H. (2010) A new locus for early maturity in soybean. Crop Sciences 50:524–527

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., Coupland, G. (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science 316:1030-1033

Covington, M.F., Maloof, J.N., Straume, M., Kay, S.A., Harmer, S.L. (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biol. 9:R130

Davis, S.J., Millar, A.J. (2001) Watching the hands of the Arabidopsis biological clock. Genome Biol. 2:1001–1008

Davis, S.J. (2002) Photoperiodism: The coincidental perception of the season. Curr. Biol. 12: R841–R843

De Lucia, F., Crevillen, P., Jones, A., Greb, T., Dean, C. (2008) A PHD-Polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proc. Natl Acad. Sci. USA 105, 16831–16836

Devlin P.F., Kay S.A. (2000) Cryptochromes are required for phytochrome signaling to the circadian clock but not for rhythmicity. The Plant Cell;12:2499-2509

Dixon, L.E., Knox, K., Kozma-Bognar, L., Southern, M.M., Pokhilko, A., Millar, A.J. (2011) Temporal repression of core circadian genes is mediated through EARLY FLOWERING 3 in Arabidopsis. Curr. Biol. 21: 120–125

Domagalska, M., Sarnowska, E., Nagy, F., Davis, S. Genetic analyses of interactions among gibberellin, abscisic acid, and brassinosteroids in the control of flowering time in *Arabidopsis thaliana*. PLoS ONE 5:e14012

Doyle, M., Davis, J., Bastow, R., McWatters, H., Kozma-Bognar, L., Nagy, F., Millar, A., Amasino, R. (2002) The *ELF4* gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. *Nature* 419:74-77

Edwards, D. and Batley, J. (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnology Journal 8:2-9

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., deWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S. (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133-138

Fornara F., Panigrahi K., Gissot L., Sauerbrunn N., Ruhl M., Jarillo J., Coupland G. (2009) Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Dev Cell 17:75–86

Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., Coupland, G., Putterill, J. (1999) GIGANTEA: A circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. EMBO J. 18: 4679–4688

Franklin, K., Praekelt, U., Stoddart, W., Billingham, O., Halliday, K., Whitelam, G. (2003) Phytochromes B, D, and E act redundantly to control multiple physiological responses in *Arabidopsis*. Plant Physiol. 131(3):1340-1346

Franklin K., Davis S.J., Stoddart W., Vierstra R.D., Whitelam G. (2003a) Mutant analyses define multiple roles for phytochrome C in Arabidopsis thaliana photomorphogenesis. The Plant Cell;15:1981-1989

Franklin, K., Whitelam G. (2004) Light signals, phytochromes and cross-talk with other environmental cues. J. Exp. Bot. 55:271–276

Garner, W., Allard, H. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. J. Agric. Res. 8:553–606

Garner, W., Allard, H. (1923). Further studies in photoperiodism, the response of the plant to relative length of day and night. J. Agric. Res. 23:871–920

Han, T., Wu, C., Tong, Z., Mentreddy, R., Tan, K., Gai, J. (2006) Postflowering photoperiod regulates vegetative growth and reproductive development of soybean. Env. and Exp. Botany 55:120-129

Hanano S, Stracke, R., Jakoby, M., Merkle, T., Domagalska, M., Weisshaar, B., Davis, S. (2008) A systematic survey in Arabidopsis thaliana of transcription factors that modulate circadian parameters. BMC Genom 9:182

Harris, T., Buzby, P., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., DiMeo, J., Efcavitch, J.W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S., Steinmann, K., Thayer, E., Tyurina, A., Ward, R., Weiss, H., Xie, Z. Single-molecule DNA sequencing of a viral genome. Science 320:106-109

Hartmann U., Hohmann S., Nettesheim K., Wisman E., Saedler H., Huijser P. (2000) Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. Plant Jour. 21:351–360

Hayama, R., Izawa, T., Shimamoto, K. (2002) Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. Plant Cell Physiol. 43:494-504

Hayama, R., Coupland, G. (2003) Shedding light on the circadian clock and the photoperiodic control of flowering. Current Opinion in Plant Biology 6:13-19

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., Shimamoto, K. (2003) Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 422:719-722

Hayama, R., Agashe, B., Luley, E., King, R., Coupland, G. (2007) A circadian rhythm set by dusk determines the expression of *FT* homologs and the short-day photoperiodic flowering response in Pharbitis. The Plant Cell 19:2988-3000

Hazen, S., Schultz, T., Pruneda-Paz, J., Borevitz, J., Ecker, J., Kay, S. (2005) *LUX ARRHYTHMO* encodes a MYB domain protein essential for circadian rhythms. PNAS 102:10387-10392

Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J., Vardy, M.E., Ellis, N., Pio Beltran, J., Rameau, C., Weller, J.L. (2005) Conservation of Arabidopsis flowering genes in model legumes. Plant Physiol. 137(4):1420-1434

Helfer, A., Nusinow, D.A., Chow, B.Y., Gehrke, A.R., Bulyk, M.L., Kay, S.A. (2011) *LUX ARRHYTHMO* encodes a nighttime repressor of circadian gene expression in the Arabidopsis core clock. Curr. Biol. 21: 126–133

Hepworth, S., Valverde, F., Ravenscroft, D., Mouradov, A., Coupland, G. (2002) Antagonistic regulation of flowering-time gene *SOC1* by CONSTANS and FLC via separate promoter motifs. The EMBO Jour. 21:4327-4337

Herrero, E., Kolmos, E., Bujdoso, N., Yuan, Y., Wang, M., Berns, M., Uhlworm, H., Coupland, G., Saini, R., Jaskolski, M., Webb, A., Goncalves, J., Davis, S. (2012) EARLY FLOWERING4 recruitment of EARLY FLOWERING3 in the nucleus sustains the *Arabidopsis* circadian clock. The Plant Cell 24:428-443

Hicks, K., Millar, A., Carre, I., Somers, D., Straume, M., Meeks-Wagner, D., Kay, S. (1996) Conditional circadian dysfunction of the Arabidopsis early-flowering 3 mutant. Science 274:790-792

Hicks, K., Albertson, T., Wagner, D. (2001) *EARLY FLOWERING3* encodes a novel protein that regulates circadian clock function and flowering in Arabidopsis. The Plant Cell 13:1281-1292

Hoecker, U., Quail, P. (2001) The phytochrome A-specific signaling intermediate SPA1 interacts directly with COP1, a constitutive repressor of light signaling in *Arabidopsis*. J Biol Chem 276:38173–38178

Huang, G., Ma, J., Han, Y., Chen, X., Fu, Y. (2011) Cloning and expression analysis of the soybean *CO*-like gene *GmCOL9*. Plant Mol. Biol. Rep. 29:352-359

Hughes, R., Vrana, J., Song, J., Tucker, C. (2012) A light-dependent, dark-promoted interaction between *Arabidopsis* cryptochrome 1 and phytochrome B. JBC M112.360545 <http://www.jbc.org/cgi/doi/10.1074/jbc.M112.360545>

Imaizumi, T., Schultz, T., Harmon, F., Ho, L., Kay, S. (2005) FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. Science 309:293–297

Ishikawa, R., Aoki, M., Kurotani, K., Yokoi, S., Shinomura, T., Takano, M., Shimamoto, K. (2011) Phytochrome B regulates Heading date 1 (Hd1)-mediated expression of rice florigen Hd3a and critical day length in rice. Mol. Genet. Genomics 285:461-470

Itoh, H., Nonoue, Y., Yano, M., Izawa, T. (2010) A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. Nat. Genet. 42:635-638

Jang, S., Marchal, V., Panigrahi, K., Wenkel, S., Soppe, W., Deng, X., Valverde, F., Coupland, G. (2008) *Arabidopsis* COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. EMBO J 27:1277–1288

Jiang, Y., Han, Z., Zhang, M. (2011) Expression profiles of a *CONSTANS* homologue *GmCOL11* in *Glycine max*. Russian Journal of Plant Physiology 58:928-935

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., Dean, C. (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science 290:344–347

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences 8:275-282

Jones, M. (2009) Entrainment of the *Arabidopsis* circadian clock. J. Plant Biol. 52:202-209

Jung, J., Seo, P., Ahn, J., Park, C. (2012) The Arabidopsis RNA-binding protein FCA regulates microRNA172 processing in thermosensory flowering. Journal of Biological Chemistry 287:16007-16016

Kiba, T., Henriques, R., Sakakibara, H., Chua, N. (2007) Targeted degradation of PSEUDO-RESPONSE REGULATOR5 by an SCF^{ZTL} complex regulates clock function and photomorphogenesis in *Arabidopsis thaliana*. The Plant Cell 19:2516-2530

Kim, W., Fujiwara, S., Suh, S., Kim, J., Kim, Y., Han, L., David, K., Putterill, J., Nam, H., Somers, D. (2007) ZEITLUPE is a circadian photoreceptoer stabilized by GIGANTEA in blue light. Nature 449:356-360

Kolmos E., Davis, S.J. (2007). *ELF4* as a central gene in the circadian clock. Plant Signal. Behav. 2: 370–372

Kolmos, E., Nowak, M., Werner, M., Fischer, K., Schwarz, G., Mathews, S., Schoof, H., Nagy, F., Bujnicki, J.M., Davis, S.J. (2009) Integrating ELF4 into the circadian system through combined structural and functional studies. Front. Life Sci. (formerly HFSP J.) 3: 350–366

Kolmos, E., Herrero, E., Bujdoso, N., Millar, A.J., Tóth, R., Gyula, P., Nagy, F., Davis, S.J. (2011) A reduced-function allele reveals that EARLY FLOWERING3 repressive action on the circadian clock is modulated by phytochrome signals in Arabidopsis. Plant Cell 23: 3230–3246

Komiya, R., Ikegami, A., Tamaki, S., Yokoi, S., Shimamoto, K. (2008) Hd3a and RFT1 are essential for flowering in rice. Development 135:767-774

Komiya, R., Yokoi, S., Shimamoto, K. (2009) A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice. Development 136:3443-3450

Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B., Watanabe, S., Yamada, T., Tabata, S., Kanazawa, A., Harada, K., Abe, J. (2010) Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. Plant Physiology 154:1220-1231

Koornneef, M., Hanhart, C., van der Veen, J. (1991) A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. Mol Gen Genet 229:57–66

Koornneef, M., Alonso-Blanco, C., Vries, H.B.-D., Hanhart, C.J., and Peeters, A.J.M. (1998). Genetic interactions among late-flowering mutants of Arabidopsis. Genetics 148, 885–892

Laibach, F. (1943) Arabidopsis thaliana (L.) Heynh. als Objekt fu¨r genetische und entwicklungsphysiologische Untersuchun- gen. Bot Arch 44:439–455

Langridge, J. (1957) Effect of day-length and gibberellic acid on the flowering of Arabidopsis. Nature 180:36–37

Laubinger, S., Hoecker, U. (2003) The SPA1-like proteins SPA3 and SPA4 repress photomorphogenesis in the light. Plant J 35:373–385

Laubinger, S., Marchal, V., Le Gourrierec, J., Gentilhomme, J., Wenkel, S., Adrian, J., Jang, S., Kulajta, C., Braun, H., Coupland, G., Hoecker, U. (2006) *Arabidopsis* SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. Development 133:3213–3222

Lee, J., Yoo, S., Park, S., Hwang, I., Lee, J., Ahn, J. (2007) Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes Dev 21:397–402

Liu, L., Zhang, Y., Li, Q., Sang, Y., Mao, J., Lian, H., Wang, L., Yang, H. (2008) COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in Arabidopsis. Plant Cell 20:292–306

Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., Abe, J. (2008a) Genetic redundancy in soybean photoresponses associated with duplication of the *Phytochrome A* gene. Genetics 180:995-1007

Liu, H., Wang, H., Gao, P., Xu, J., Xu, T., Wang, J., Wang, B., Lin, C., Fu, Y. (2009) Analysis of clock gene homologs using unifoliates as target organs in soybean (*Glycine max*). Journal of Plant Physiology 166:278-289

Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., Nagamatsu, A., Arai, M., Yamada, T., Kitamura, K., Masuta, C., Harada, K., Abe, J. (2010) The soybean stem growth habit gene *Dt1* is an ortholog of Arabidopsis *TERMINAL FLOWER1*. Plant Physiology 153:198-210

Liu, L., Ma, J., Han, Y., Chen, X., Fu, Y. (2011) The isolation and analysis of a soybean *CO* homologue *GmCOL10*. Russian Journal of Plant Physiology 58:330-336

Locke, J., Kozma-Bognar, L., Gould, P., Feher, B., Kevei, E., Nagy, F., Turner, M., Hall, A., Millar, A. (2006) Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. Mol. Syst. Biol. 2:59

Mardis, E.R. (2008) Next-generation DNA sequencing methods. Annu. Rev. Genom. Human Genet. 9:387-402

Mas, P., Devlin, P., Panda, S., Kay, S. (2000) Functional interaction of phytochrome B and cryptochrome 2. Nature 408:207-211

Mas, P., Kim, W., Somers, D., Kay, S. (2003) Targeted degradation of TOC1 by ZTL modulates circadian function in *Arabidopsis thaliana*. Nature 426:567-570

Matsumura, H., Kitajima, H., Akada, S., Abe, J., Minaka, N., Takahashi, R. Molecular cloning and linkage mapping of cryptochrome multigene family in soybean. The Plant Genome 2:271-281

McBlain, B., Bernard, R.L. (1987) A new genetic effecting the time of flowering and maturity in soybean. Jour. Hereditary 78:160–162

McWatters, H., Kolmos, E., Hall, A., Doyle, M., Amasino, R., Gyula, P., Nagy, F., Millar, A., Davis, S. (2007) *ELF4* is required for oscillatory properties of the circadian clock. Plant Physiology 144:391-401

Michaels, S., Amasino, R. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11:949-956

Mockler, T., Guo, H., Yang, H., Duong, H., Lin, C. (1999) Antagonistic actions of *Arabidopsis* cryptochromes and phytochrome B in the regulation of floral induction. Development 126:2073-2082

Mutasa-Gottgens, E., Hedden, P. (2009) Gibberellin as a factor in floral regulatory networks. J. Exp. Bot. 60:1979-1989

Nakamichi, N., Kita, M., Ito, S., Yamashino, T., Mizuno, T. (2005) PSEUDO-RESPONSE REGULATORS, PRR9, PRR7, and PRR5, together play essential roles close to the circadian clock of *Arabidopsis thaliana*. Plant Cell Physiol. 46:686-698

Nakamichi, N., Kiba, T., Henriques, R., Mizuno, T., Chua, N., Sakakibara, H. (2010) PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the Arabidopsis circadian clock. Plant Cell 22:594-605

Nakamichi, N. (2011) Molecular mechanisms underlying the *Arabidospsis* circadian clock. Plant Cell Physiol. 52:1709-1718

Onai, K., Ishiura, M. (2005) *PHYTOCLOCK1* encoding a novel GARP protein essential for the *Arabidopsis* circadian clock. Genes to Cells 10:963-972

Parks, B., Quail, P. (1993) hy8, a new class of Arabidopsis long hypocotyl mutants deficient in functional Phytochrome A. The Plant Cell 5:39-48

Pittendrigh, C.S., and Minis, D.H. (1964). The entrainment of circadian oscillations by light and their role as photoperiodic clocks. Am. Nat. 108, 261–293

Portoles, S., Mas, P. (2010) The functional interplay between protein kinase CK2 and CCA1 transcriptional activity is essential for clock temperature compensation in Arabidopsis. PLoS Genetics 6:e1001201

Purvis, O.N., Gregory, F.G. (1952) The reversibility by high tem- perature of the vernalised condition in Petkus winter rye. Ann Bot 16:1–21

Putterill, J., Robson, F., Lee, K., Simon, R., Coupland, G. (2004) The CONSTANS gene of arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80:847-857

Quecini, V., Zucchi, M., Baldin, J., Vello, N. (2007) Identification of soybean genes involved in circadian clock mechanism and photoperiodic control of flowering time by *in silico* analyses. Journal of Integrative Plant Biology 49:1640-1653

Redei G.P. (1962) Supervital mutants of Arabidopsis. Genetics 47:443–460

Reed, J., Nagatani, A., Elich, T., Fagan, M., Chory, J. (1994) Phytochrome A and Phytochrome B have overlapping but distinct functions in Arabidopsis development. Plant Physiology 104:1139-1149

Rehm, G., Randall, G., Lamb, J., Eliason, R. (2006) Fertilizing Corn in Minnesota. Publication #FO-0379O, University of Minnesota Extension

Samach, A., Onuichi, H., Gold, S., Ditta, G., Schwarz-Sommer, Z., Yanofsky, M., Coupland, G. (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science 288, 1613-1616

Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. PNAS 74:5463-5467

Sawa M, Nusinow D, Kay S, Imaizumi T (2007) FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science 318:261–265

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carré, I.A., Coupland, G. (1998) The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell 93: 1219–1229

Schwab, R. (2012) The roles of miR156 and miR172 in phase change regulation. Signaling and Communication in Plants 15:49-68

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. (2010) Genome sequence of the palaeopolyploid soybean. Nature 465:178-183

Shin, J., Lee, S. (2012) Molecular markers for the *E2* and *E3* genes controlling flowering and maturity in soybean. Mol. Breeding DOI 10.1007/s11032-012-9743-6

Simpson, G. (2005) The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. Current Opinion in Plant Biology 7:570-574

Smith, H., Whitelam, G.C. (1997) The shade avoidance syndrome: multiple responses mediated by multiple phytochromes. Plant Cell Environ 20(6):840–844

Somers, D.E., Devlin, P.F., Kay, S.A. (1998) Phytochromes and cryptochromes in the entrainment of the Arabidopsis circadian clock. Science 282:1488–1490

Srikanth, A., Schmid, M. (2011) Regulation of flowering time: all roads lead to Rome. Cell. Mol. Life Sci. 68:2013-2037

Stewart, D., Cober, E., Bernard, R. (2003) Modeling genetic effects on the photothermal response of soybean phonological development. Agronomy Journal 95:65-70

Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., Coupland, G. (2001) *CONSTANS* mediates between the circadian clock the control of flowering in *Arabidopsis*. Nature 410:1116-1120

Sun, H., Jia, Z., Cao, D., Jiang, B., Wu, C., Hou, W., Liu, Y., Fei, Z., Zhao, D., Han, T. (2011) *GmFT2a*, a soybean homolog of *FLOWERING LOCUST T*, is involved in flowering transition and maintenance. PLoS ONE 6:e29238

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 462, 799–802

Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S., Shimamoto, K. (2007) Hd3a protein is a mobile flowering signal in rice. Science 316:1033-1036

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution (In Press)

Thakare, D., Kumudini, S., Dinkins, R. (2010) Expression of flowering-time genes in soybean E1 near-isogenic lines under short and long day conditions. Planta 231:951-963

Thakare, D., Kumudini, S., Dinkins, R. (2011) The alleles at the E1 locus impact the expression of two soybean *FT*-like genes shown to induce flowering in Arabidopsis. Planta 234:933-943

Thines, B., Harmon, F. (2010) Ambient temperature response establishes ELF3 as a required component of the core *Arabidopsis* circadian clock. PNAS 107:3257-3262

Tóth R, Kevei E, Hall A, Millar A, Nagy F, Kozma-Bognar L (2001) Circadian clock-regulated expression of phytochrome and cryptochrome genes in *Arabidopsis*. Plant Physiol 127:1607–1616

Tsuchiya T, Ishiguri Y (1981) Role of the quality of light on the photoperiodic flowering response in four latitudinal ecotypes of Chenopodium rubrum. Plant Cell Physiol 22:525–532

Turck, F., Fornara, F., Coupland, G. (2008) Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annu. Rev. Plant Biol. 59:573-594

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. (2004) Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* 303:1003–1006

Wang, J., Czech, B., Weigel, D. (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. Cell 138:625-627

Wang, L., Fujiwara, S., Somers, D. (2010) PRR5 regulates phosphorylation, nuclear import and subnuclear localization of TOC1 in the Arabidopsis circadian clock. The EMBO Journal 29:1903-1915

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S., Kitamura, K., Harada, K. (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. Genetics 188:395-407

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C., Irzyk, G., Lupski, J., Chinault, C., Song, X., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D., Marguilies, M., Weinstock, G., Gibbs, R., Rothberg, J. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872–876

Wu, F., Zhang, X., Li, D., Fu, Y. (2011) Ectopic expression reveals a conserved *PHYB* homolog in soybean. PLoS ONE 6:e27737

Xia, Z., Zhai, H., Liu, B., Kong, F., Yuan, X., Wu, H., Cober, E., Harada, K. (2012) Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean. Plant Syst. Evol. DOI 10.1007/s00606-012-0628-2

Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., Anai, T., Sato, S., Yamazaki, T., Lu, S., Wu, H., Tabata, S., Harada, K. (2012a) Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. www.pnas.org/cgi/doi/10.1073/pnas.1117982109

Xu, J., Zhong, X., Zhang, Q., Li, H. (2010) Overexpression of the *GmGAL2* gene accelerates flowering in *Arabidopsis*. Plant Mol. Biol. Rep. 28:704-711

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., Sasaki, T. (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. The Plant Cell 12:2473-2483

Yanovsky, M.J., Kay, S.A. (2002) Molecular basis of seasonal time measurement in Arabidopsis. *Nature* 419:308–312

Yu, J.W., Rubio, V., Lee, N.Y., Bai, S., Lee, S.Y., Kim, S.S., Liu, L., Zhang, Y., Irigoyen, M.L., Sullivan, J.A., Zhang, Y., Lee, I., Xie, Q., Paek, N.C., Deng, X.W. (2008) COP1 and ELF3 control circadian function and photoperiodic flowering by regulating GI stability. Mol Cell. 2008, 32:617–30

Zeevaart, J.A.D. (1983). Gibberellins and flowering. In The Biochemistry and Physiology of Gibberellins, Vol. 2, A. Crozier, ed (New York: Praeger Scientific), pp. 333–374

Zeevaart, J.A.D. (2006) Florigen coming of age after 70 years. Plant Cell 18:1783–1789

Zeilinger M., Farre, E., Taylor, S., Kay, S., Doyle F. (2006) A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. Mol. Syst. Biol. 2:58

Zhang, Q., Li, H., Li, R., Hu, R., Fan, C., Chen, F., Wang, Z., Liu, X., Fu, Y., Lin, C. (2008) Association of the circadian rhythmic expression of GmCRY1a with a latitudinal cline in photoperiodic flowering of soybean. PNAS 105:21028-21033

Zhong, X., Dai, X., Jiaohui, X., Hanying, W., Liu, B., Li, H. (2012) Cloning and expression analysis of *GmGAL1*, *SOC1* homolog gene in soybean. Mol. Biol. Rep. 39:6967-6974

FIGURES

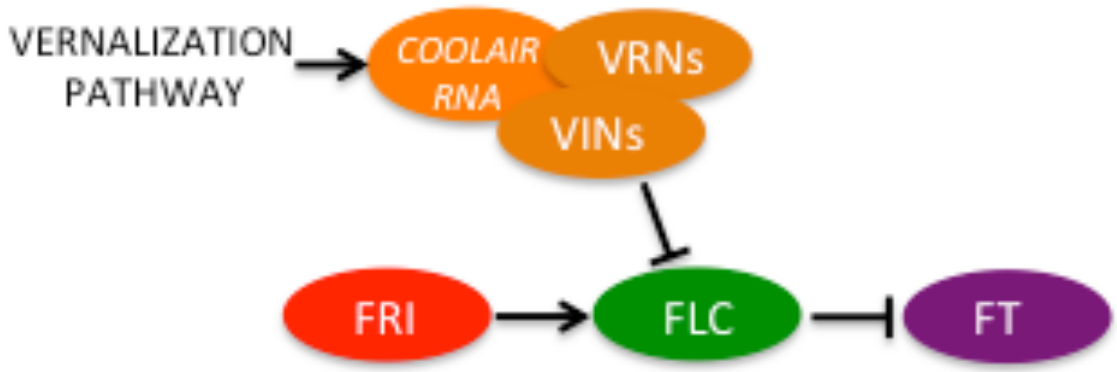

Figure 1. Illustration of the vernalization pathway in *Arabidopsis thaliana.*

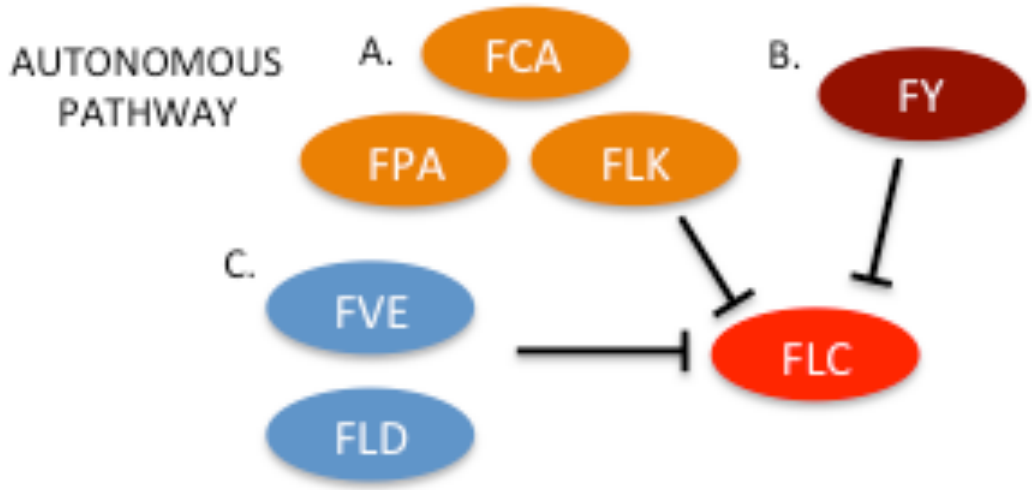

Figure 2. Illustration of the autonomous pathway in *Arabidopsis thaliana.*

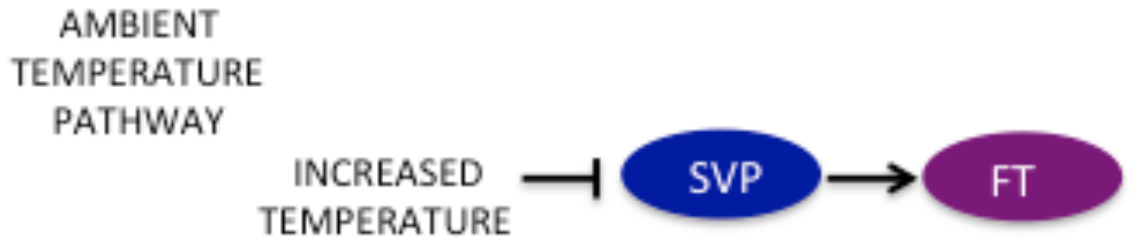

Figure 3. Illustration of the ambient temperature pathway in *Arabidopsis thaliana*.

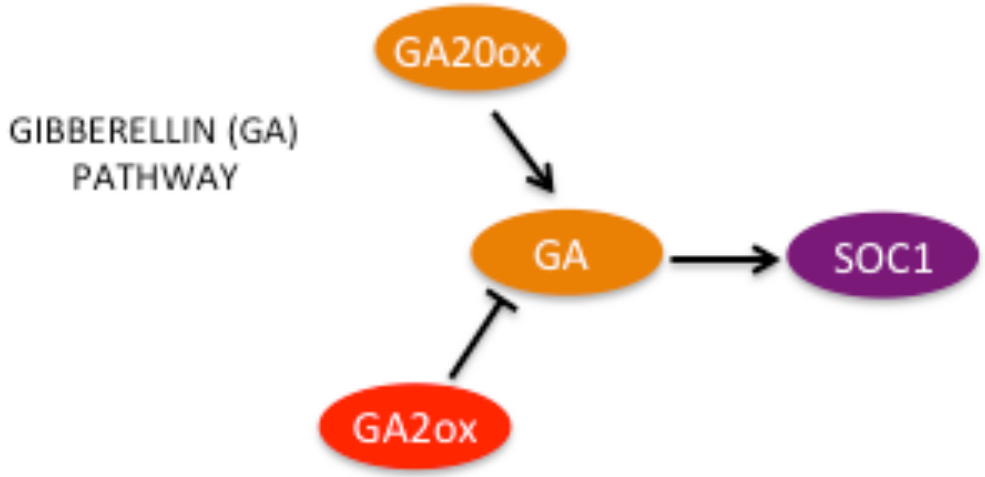

Figure 4. Illustration of the Gibberellin pathway in *Arabidopsis thaliana*.

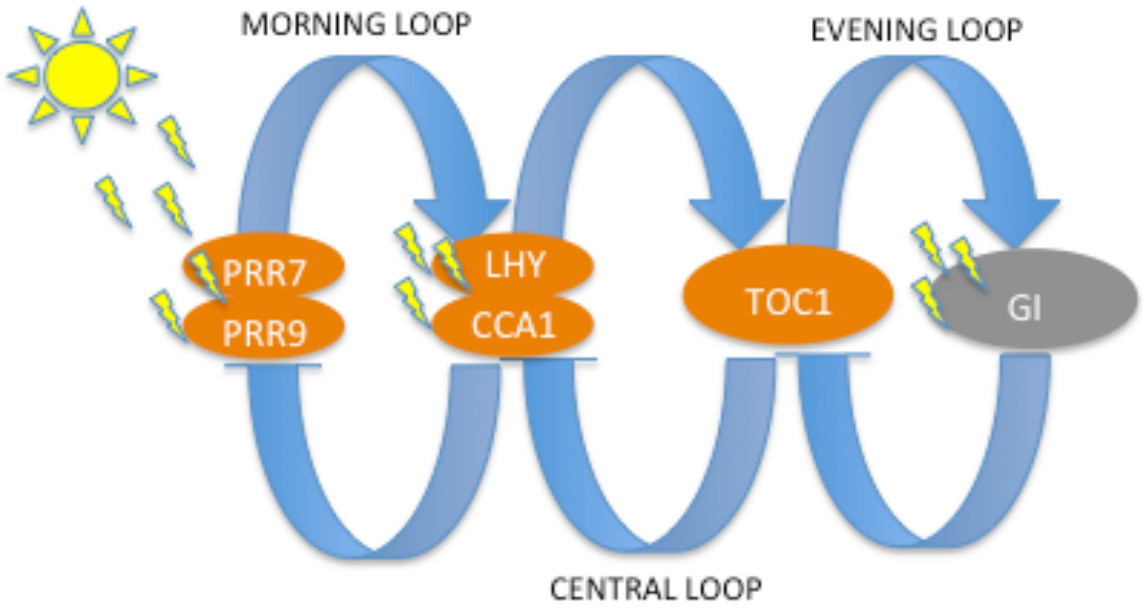

Figure 5. An illustration of the circadian clock components in *Arabidopsis thaliana* with the "sun" and "ray" symbols representing light induction.

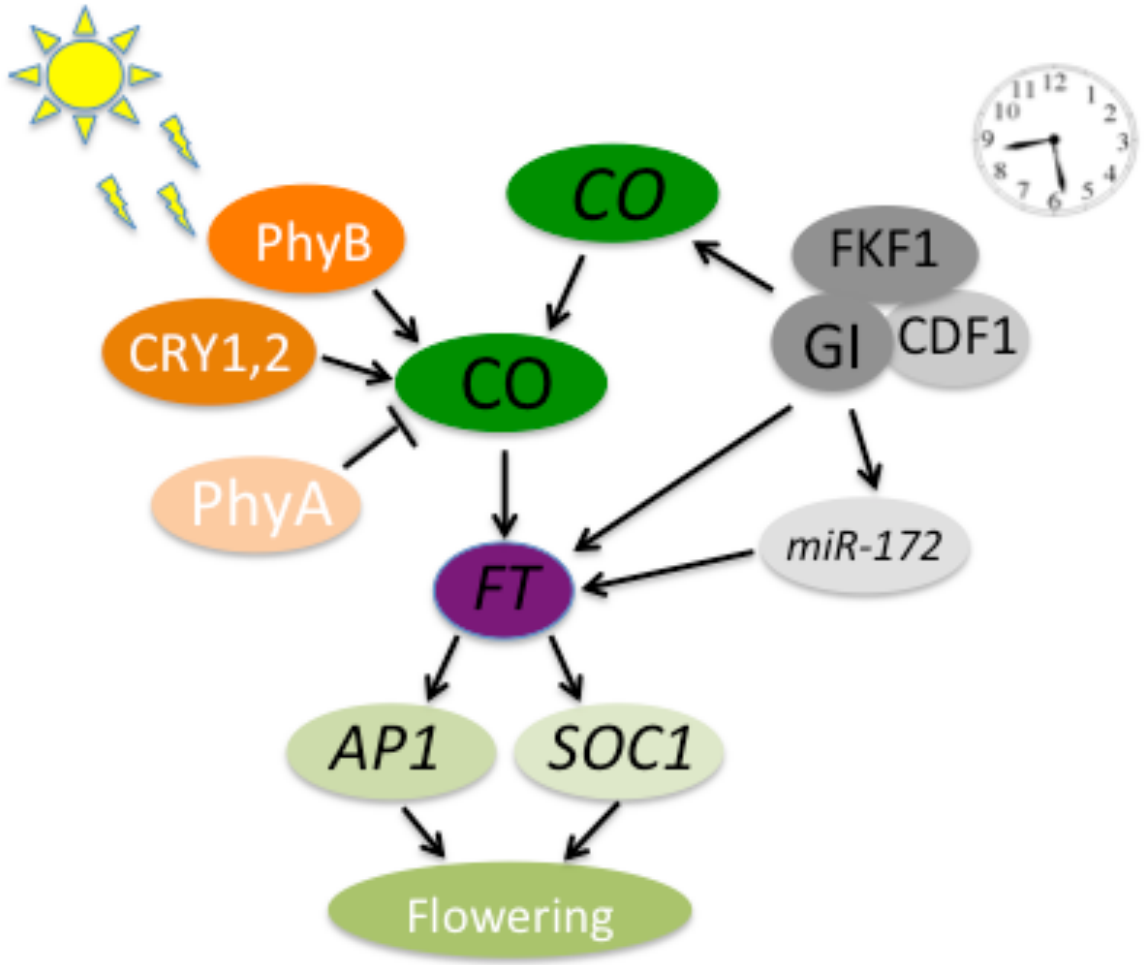

Figure 6. Simplified illustration of the photoperiod flowering pathway in *Arabidopsis thaliana*. The "clock" symbol represents circadian rhythm control of induction. The "sun" symbol represents light induction.

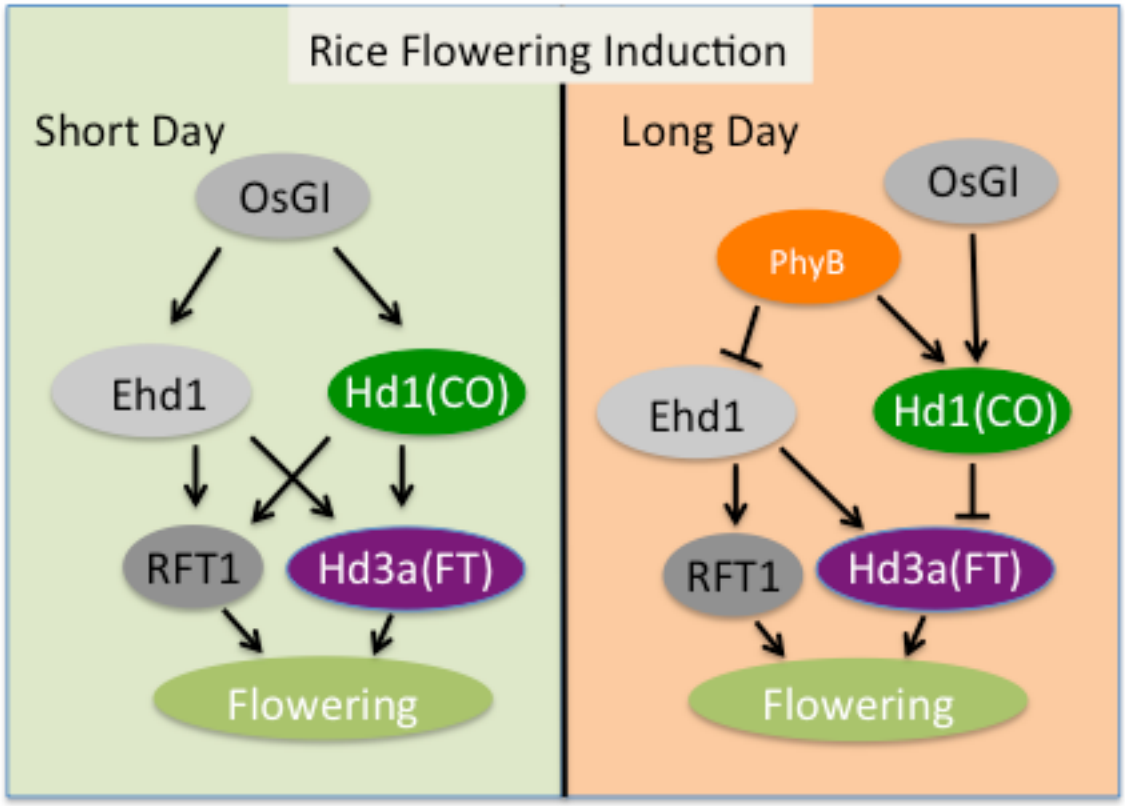

Figure 7. Simplified illustration of the photoperiod flowering pathways in *Oryza sativa* for Short Day and Long Day.

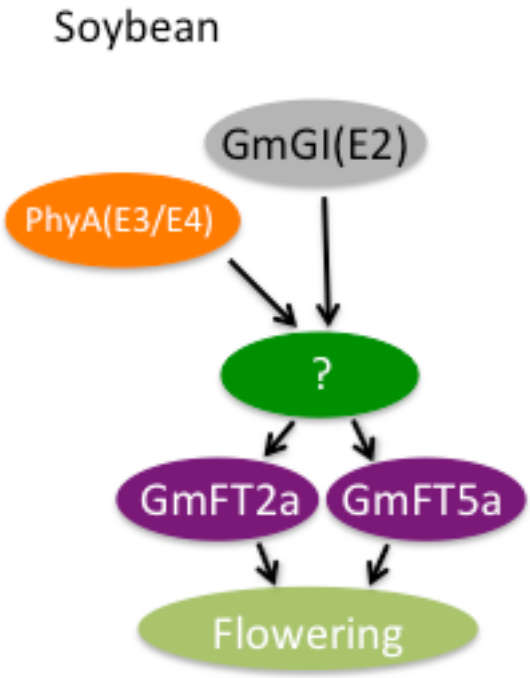

Figure 8. Simplified illustration of the photoperiod flowering pathway in *Glycine max*.

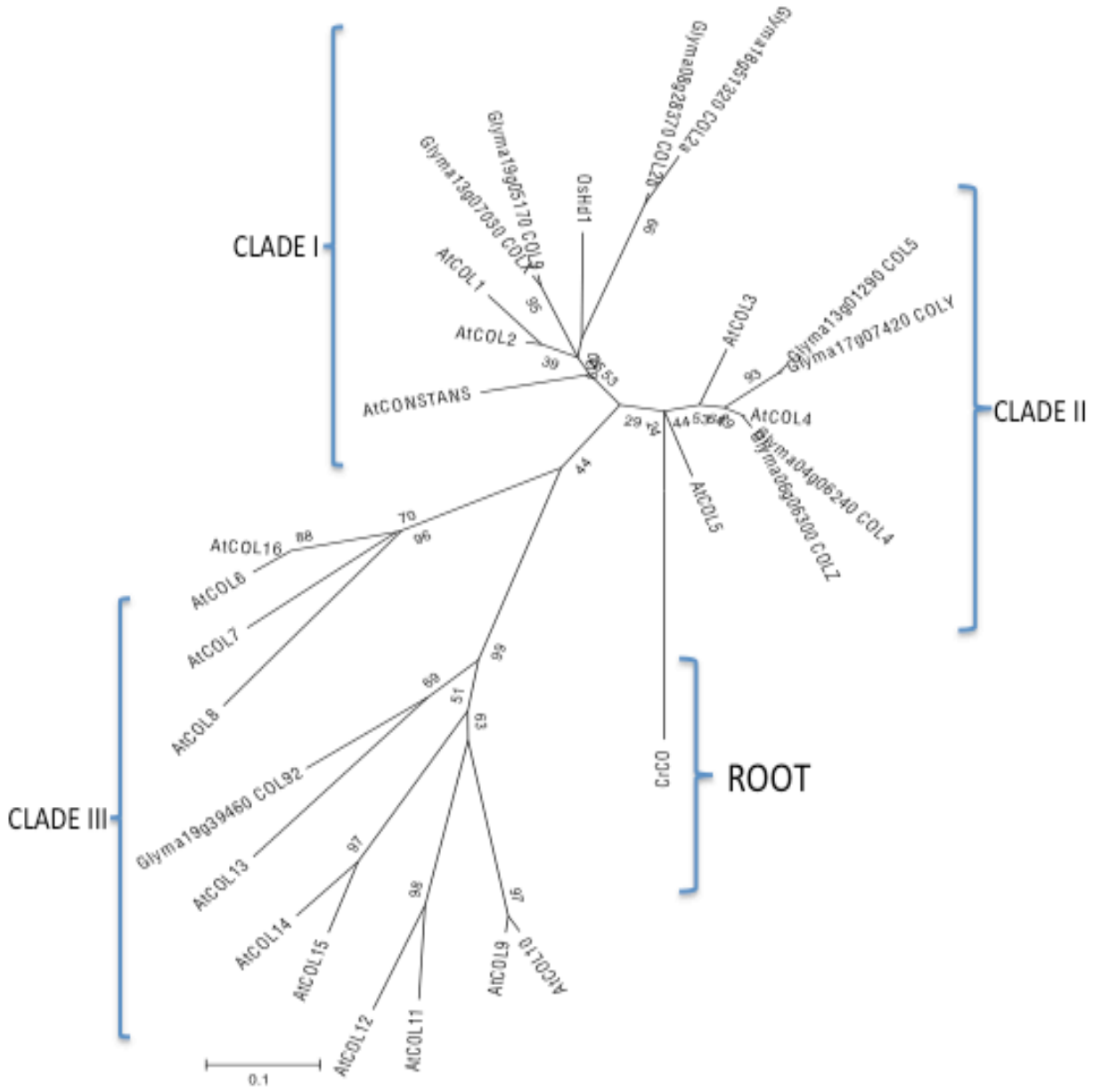

Figure 9. Phylogenetic tree containing *Arabidopsis thaliana* CONSTANS (AtCONSTANS), sixteen *A. thaliana* CONSTANS-like homologs, *Oryza sativa* Heading date 1 (OsHd1), and the nine *Glycine max* CONSTANS-like homologs. Tree was constructed using the manually edited multiple sequence alignment from MAFFT of amino acid sequences (Table 4) from phytozome.net in MEGA 5.05 software. Gene descriptions found in Table 3.

The selected *G. max* CONSTANS-like homologs appear to have a closer evolutionary relationship to *A. thaliana* than most of the *A. thaliana CONSTANS-like* genes. This suggests that CO-like proteins in soybean may be functionally analogous.

Eight of the nine CONSTANS-like homologs appear to constitute four pairs of homeologous genes, and the four pairs can be put into two groups, where the two groups indicate a closer divergent relationship. This would make sense given the two genome duplication events that have occurred in soybean.

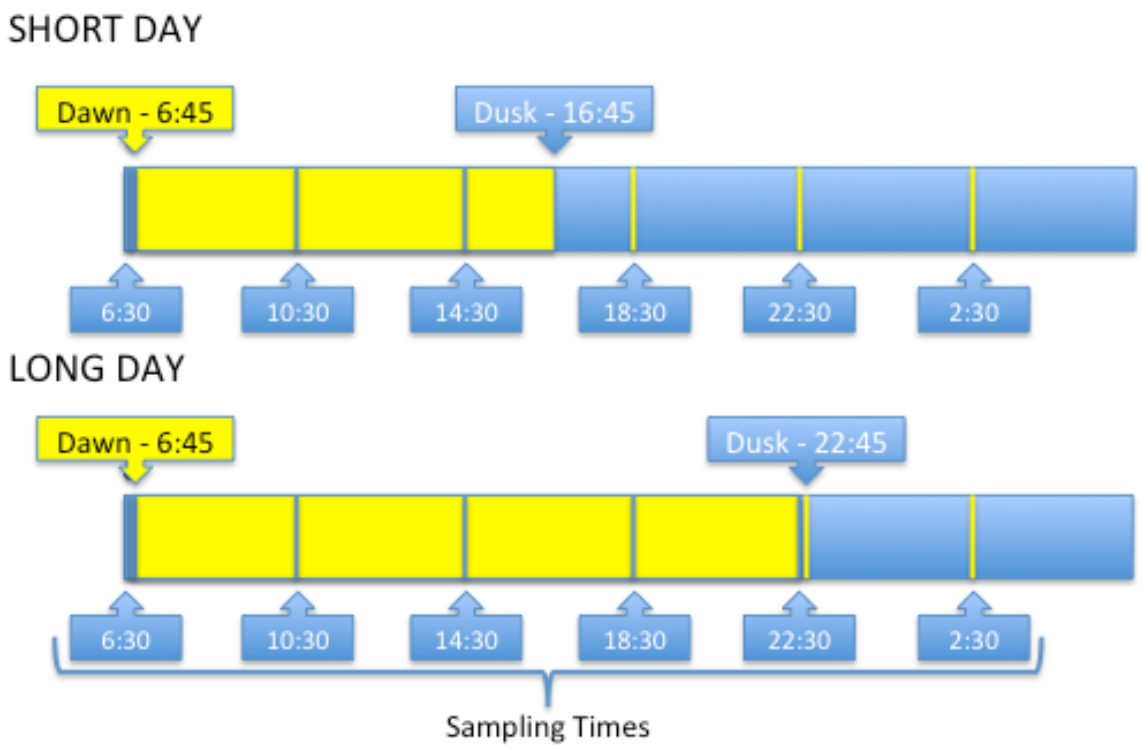

Figure 10. Day length and Sampling times. Short day (SD) with 10 hour days - dawn at 6:45 and dusk at 16:45. Long day (LD) with 16 hour days – dawn at 6:45 and dusk at 22:45. Six sampling times: 6:30, 10:30, 14:30, 18:30, 22:30, 2:30.

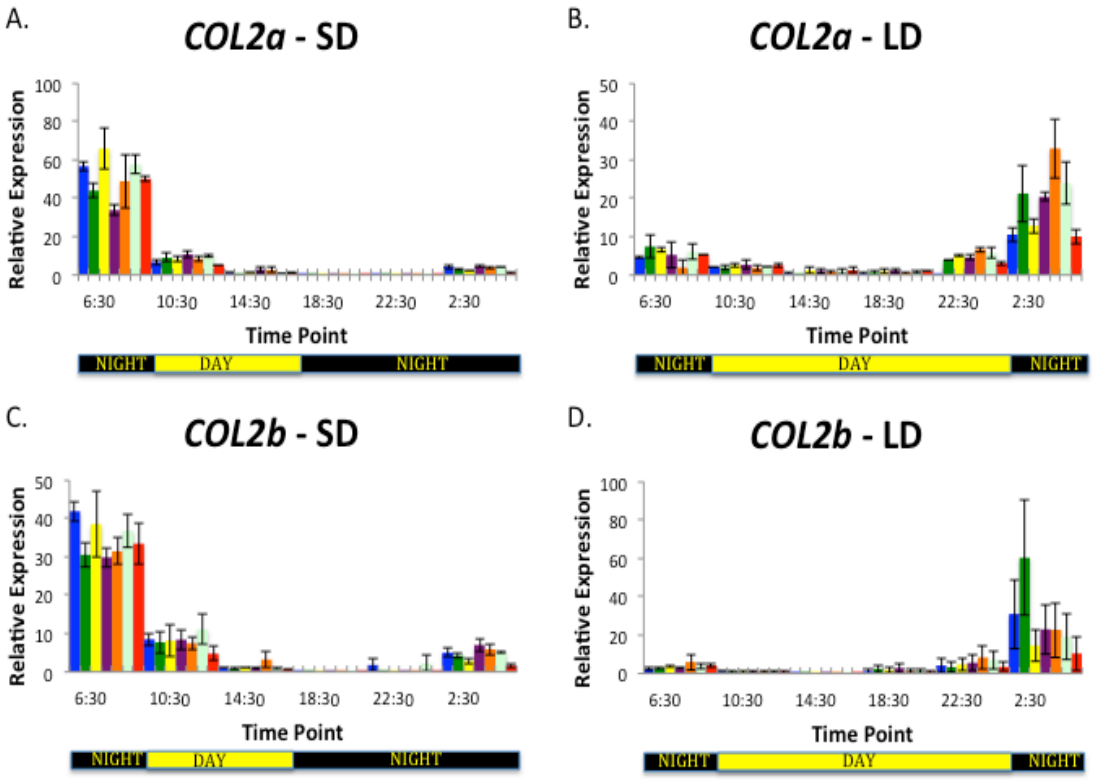

Figure 11A-D. (A) qRT-PCR results for *CONSTANS-like 2a* (*COL2a*) in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) Results for *COL2a* in Long Day (LD: dawn- 6:45/ dusk-22:45). (C) qRT-PCR results for *CONSTANS-like 2b* (*COL2b*) in Short Day. (D) Results for *COL2b* in Long Day. Relative expression (y-axis) shown at six different Time Points (x-axis). Time points sampled: 6:30, 10:30, 14:30, 18:30, 22:30, and 2:30. Light bar at the bottom indicates which Time Points fall in night and day. Standard error calculated from 3 biological replications. Data bar color reflects 7 different genotypes from Table 1.

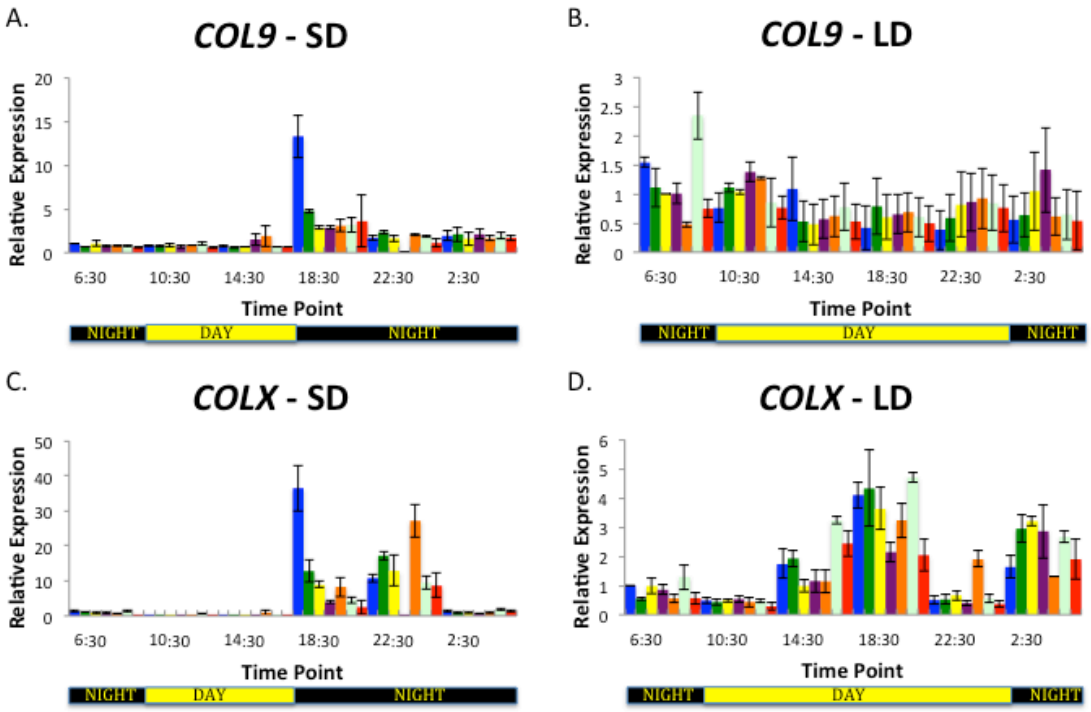

Figure 12A-D. (A) qRT-PCR results for *CONSTANS-like 9* (*COL9*) in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) Results for *COL9* in Long Day (LD: dawn- 6:45/ dusk-22:45). (C) qRT-PCR results for *CONSTANS-like X* (*COLX*) in Short Day. (D) Results for *COLX* in Long Day. Relative expression (y-axis) shown at six different Time Points (x-axis). Time points sampled: 6:30, 10:30, 14:30, 18:30, 22:30, and 2:30. Light bar at the bottom indicates which Time Points fall in night and day. Standard error calculated from 3 biological replications. Data bar color reflects 7 different genotypes from Table 1.

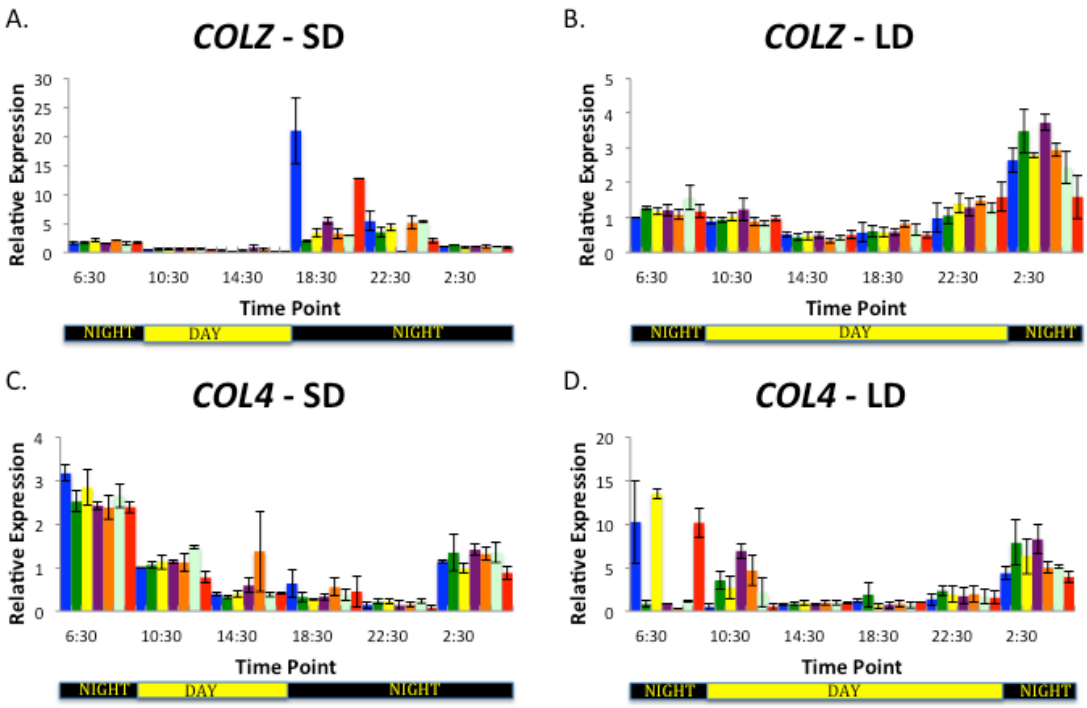

Figure 13A-D. (A) qRT-PCR results for *CONSTANS-like Z* (*COLZ*) in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) Results for *COLZ* in Long Day (LD: dawn- 6:45/ dusk-22:45). (C) qRT-PCR results for *CONSTANS-like 4* (*COL4*) in Short Day. (D) Results for *COL4* in Long Day. Relative expression (y-axis) shown at six different Time Points (x-axis). Time points sampled: 6:30, 10:30, 14:30, 18:30, 22:30, and 2:30. Light bar at the bottom indicates which Time Points fall in night and day. Standard error calculated from 3 biological replications. Data bar color reflects 7 different genotypes from Table 1.

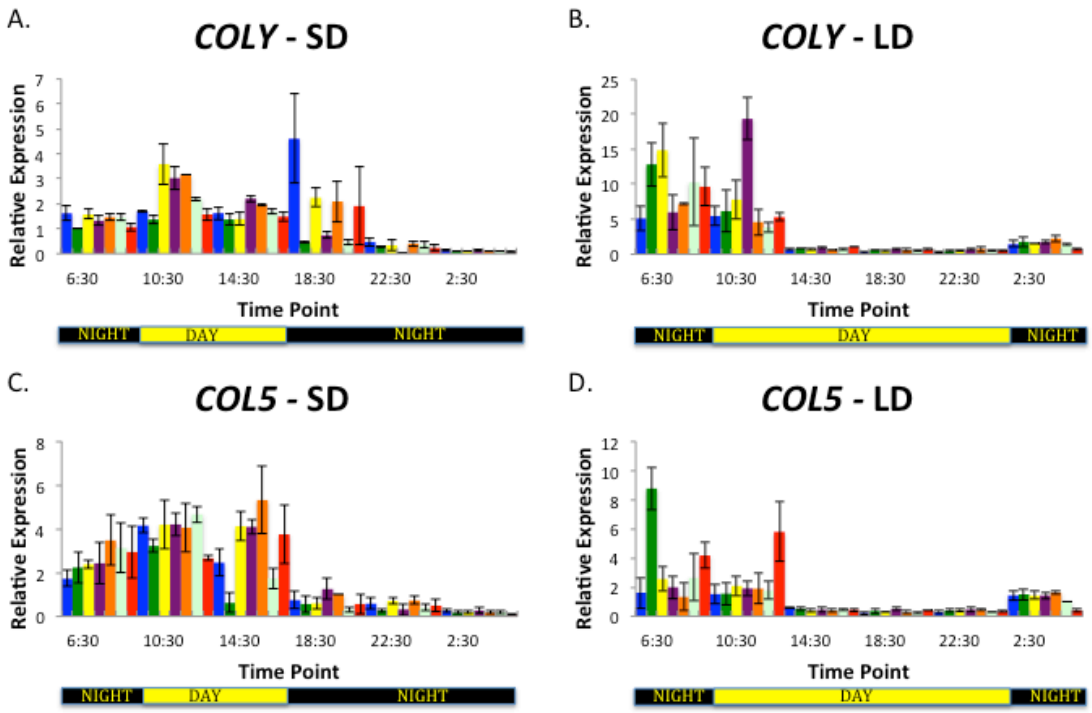

Figure 14A-D. (A) qRT-PCR results for *CONSTANS-like Y* (*COLY*) in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) Results for *COLY* in Long Day (LD: dawn- 6:45/ dusk-22:45). (C) qRT-PCR results for *CONSTANS-like 5* (*COL5*) in Short Day. (D) results for *COL5* in Long Day. Relative expression (y-axis) shown at six different Time Points (x-axis). Time points sampled: 6:30, 10:30, 14:30, 18:30, 22:30, and 2:30. Light bar at the bottom indicates which Time Points fall in night and day. Standard error calculated from 3 biological replications. Data bar color reflects 7 different genotypes from Table 1.

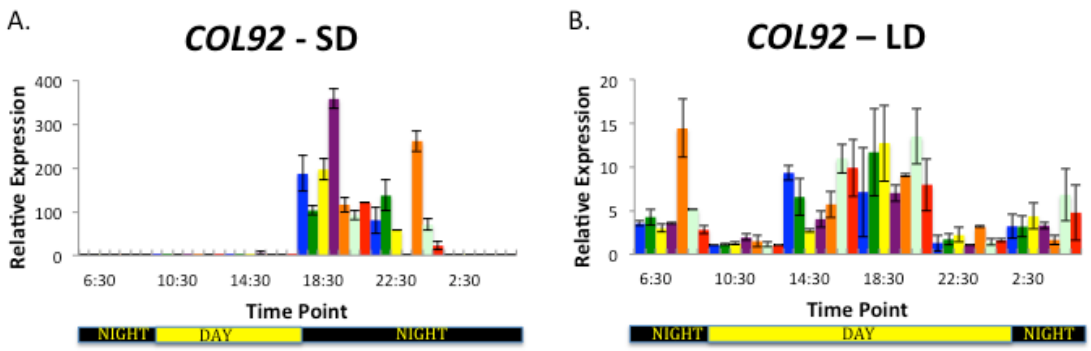

Figure 15A-B. (A) qRT-PCR results for *CONSTANS-like 92* in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) Results for *CONSTANS-like 92* in Long Day (LD: dawn- 6:45/ dusk-22:45). Relative expression (y-axis) shown at six different Time Points (x-axis). Time points sampled: 6:30, 10:30, 14:30, 18:30, 22:30, and 2:30. Light bar at the bottom indicates which Time Points fall in night and day. Standard error calculated from 3 biological replications. Data bar color reflects 7 different genotypes from Table 1.

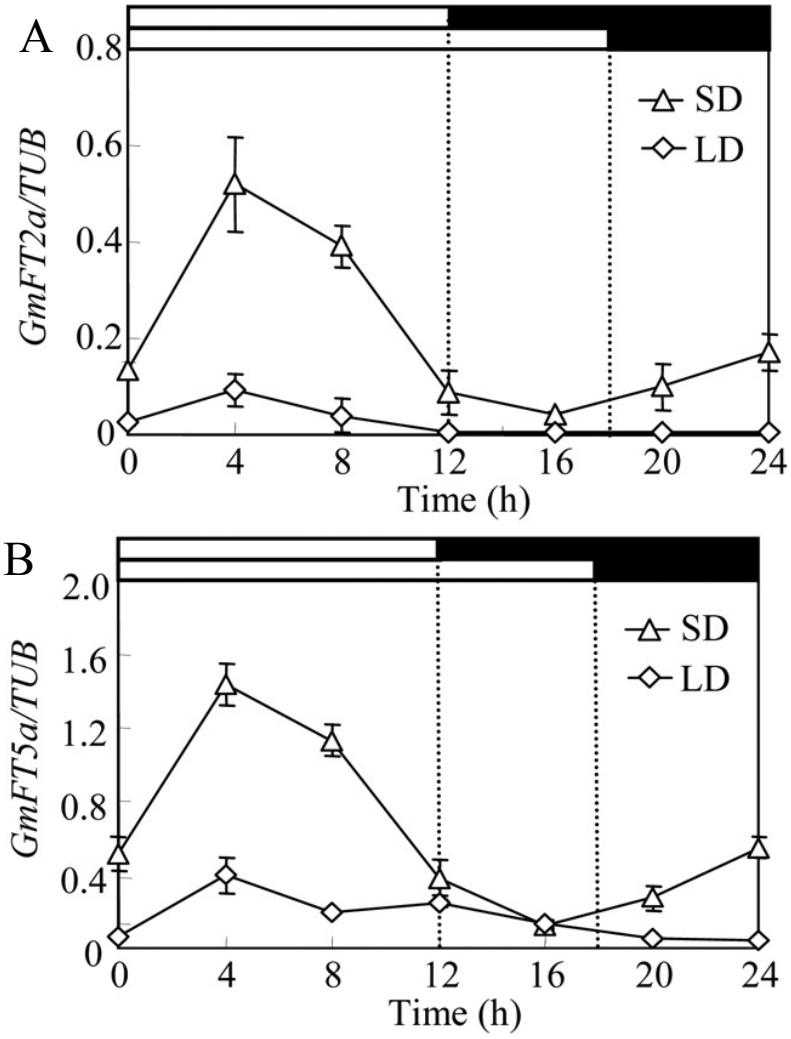

Figure 16A-B. (A) Diurnal expression of *GmFT2a* and (B) *GmFT5a* under SD and LD (Kong et al., 2010 – Modified).

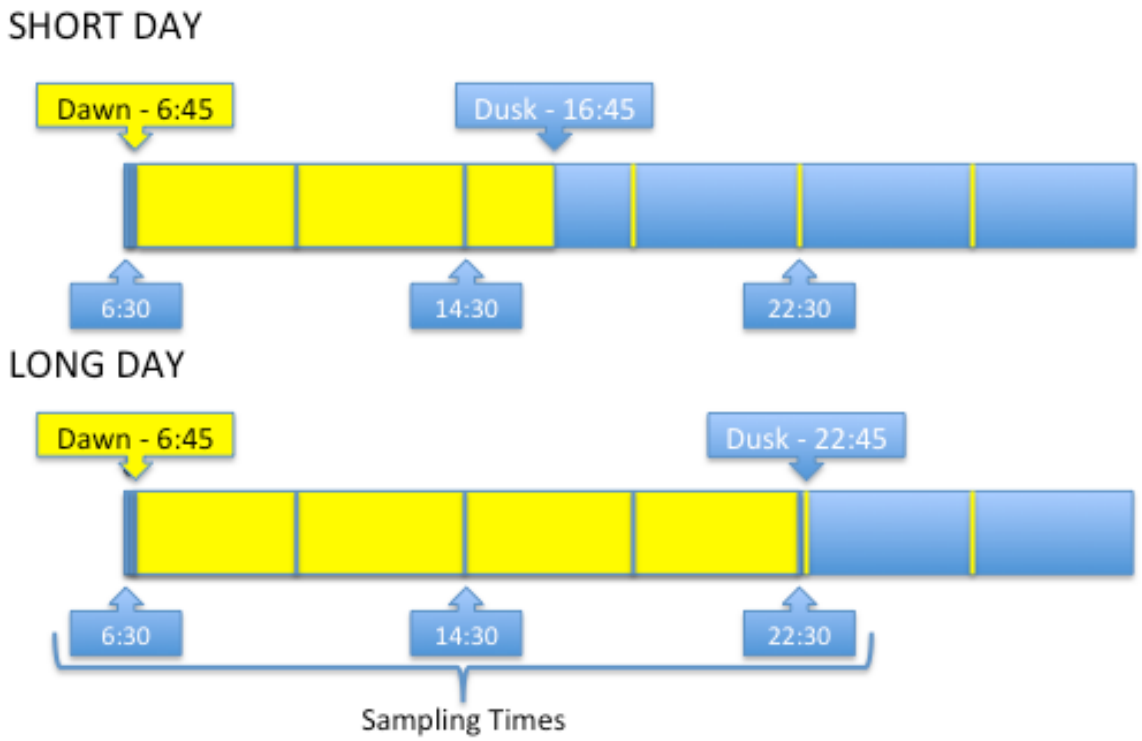

Figure 17. Sampling times used in RNA-Seq experiment. 126 samples: 2 day lengths; 3 time points; 7 genotypes; and, 3 biological replications. Results for alternative day length treatment omitted.

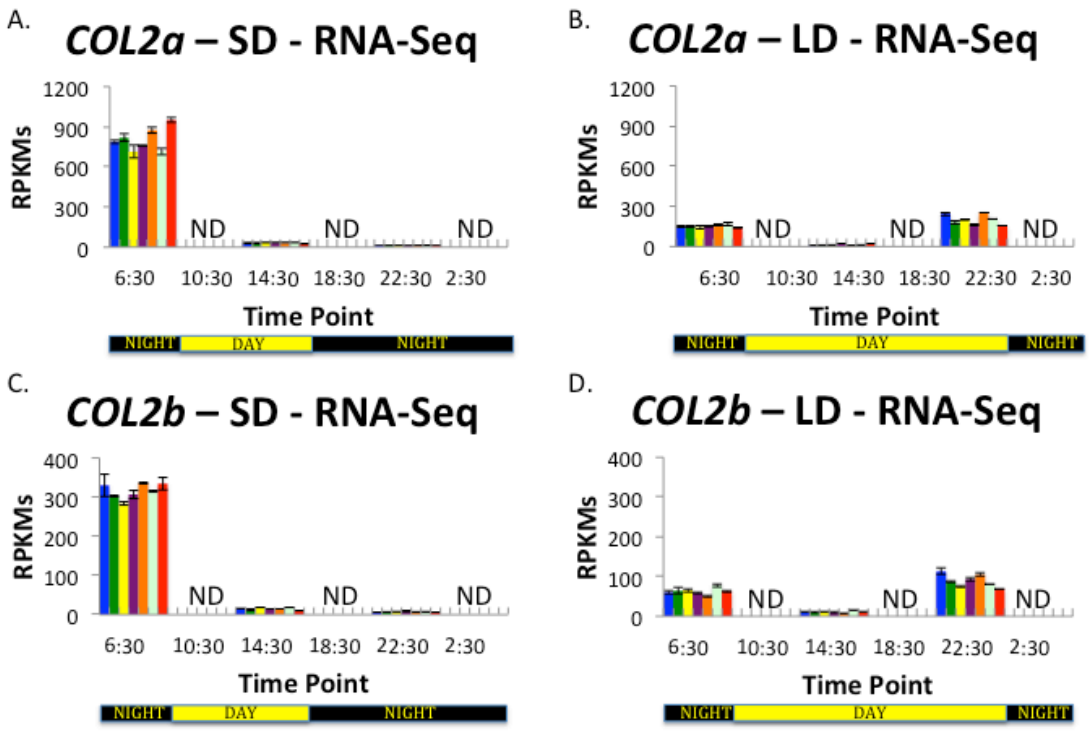

Figure 18A-D. (A) RNA-Seq results for *CONSTANS-like 2a* (*COL2a*) in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) Results for *COL2a* in Long Day (LD: dawn- 6:45/ dusk-22:45). (C) RNA-Seq results for *CONSTANS-like 2b* (*COL2b*) in SD. (D) Results for *COL2b* in LD. RPKMs, reads per kilobase of exon model per million mapped reads (y-axis) at six different Time Points (x-axis) (Mortazavi et al., 2008). Standard error calculated from 3 technical replications. Data bar color reflects genotype from Table 1. Only three Time Points sequenced [6:30; 14:30; 22:30]. ND – No Data for Time Point.

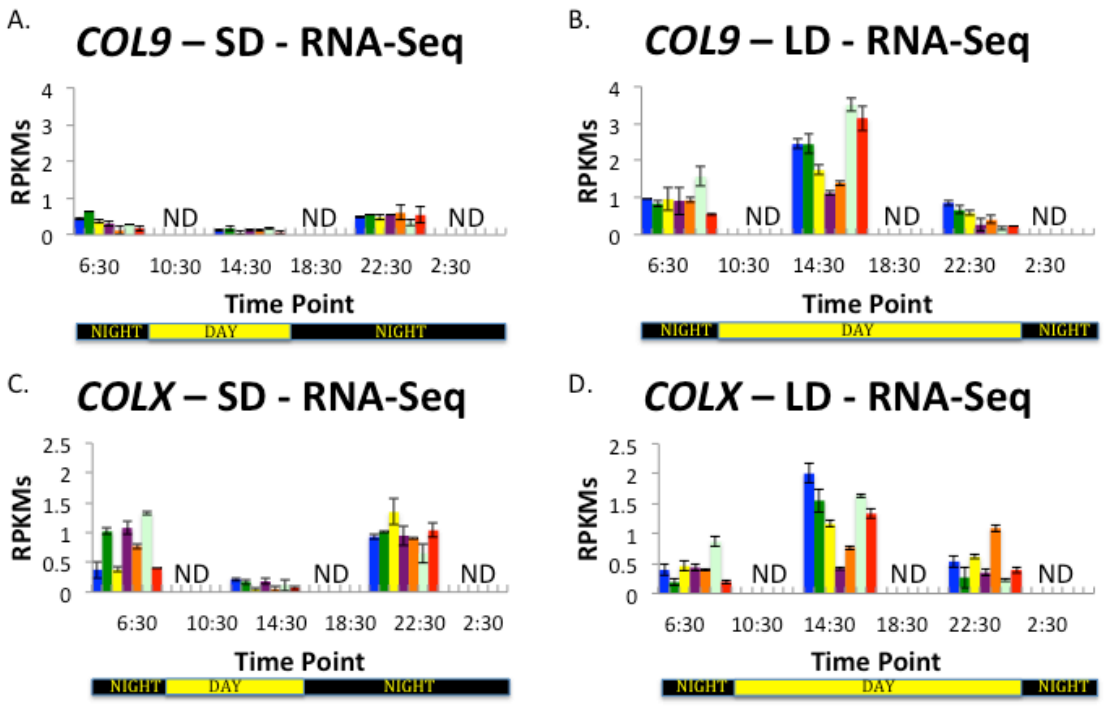

Figure 19A-D. (A) RNA-Seq results for *CONSTANS-like 9* (*COL9*) in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) Results for *COL9* in Long Day (LD: dawn- 6:45/ dusk-22:45). (C) RNA-Seq results for *CONSTANS-like X* (*COLX*) in SD. (D) Results for *COLX* in LD. RPKMs, reads per kilobase of exon model per million mapped reads (y-axis) at six different Time Points (x-axis) (Mortazavi et al., 2008). Standard error calculated from 3 technical replications. Data bar color reflects genotype from Table 1. Only three Time Points sequenced [6:30; 14:30; 22:30]. ND – No Data for Time Point.

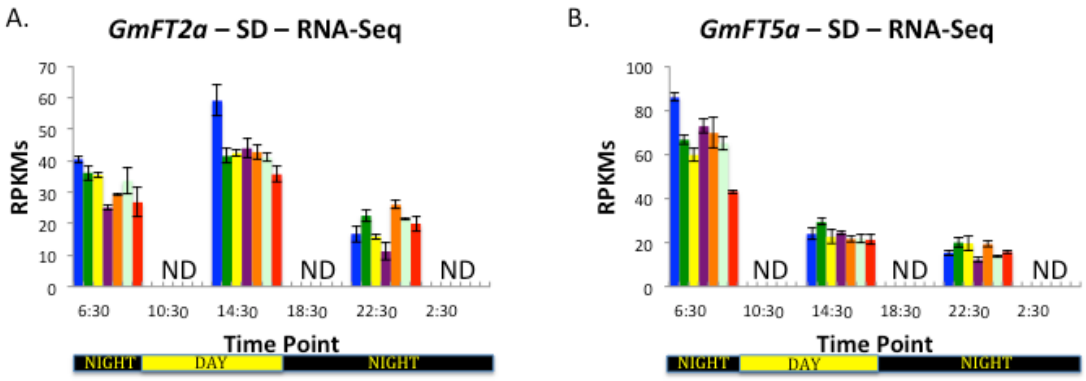

Figure 20A-B. (A) RNA-Seq results for *Glycine max FLOWERING LOCUS T 2a* (*GmFT2a*) in Short Day (SD: dawn- 6:45/ dusk-16:45). (B) RNA-Seq results for *G. max FLOWERING LOCUS T 5a* (*GmFT5a*) in Short Day (SD: dawn- 6:45/ dusk-16:45). RPKMs, reads per kilobase of exon model per million mapped reads (y-axis) at six different Time Points (x-axis) (Mortazavi et al., 2008). Standard error calculated from 3 technical replications. Data bar color reflects genotype from Table 1. Only three Time Points sequenced [6:30; 14:30; 22:30]. ND – No Data for Time Point.

TABLES

Variety	Maturity Locus (E-Locus)	Days to flower
Clark	e1E2E3E4e5E7	33
Williams 82	E2	33
L65-3366	**E1**E2E3E4e5E7	61
L66-432	e1**e2**E3E4e5E7	50
L74-441	e1E2**e3**E4e5E7	56
L92-1195	e1E2e3E4**E5**E7	36
Glycine soja	UNKNOWN	UNKNOWN

Table 1. Seven *Glycine max* genotype of the lines used in the experiment according to Maturity Loci (E-Loci). Two common North American cultivars: Clark and Williams 82; Four Near Isogenic Lines (NILs) of Clark with contrasting allele in bold; and, *Glycine soja* with unknown genotype. Days to flower are based on flowering data from 2 years (USDA - Nelson, R.).

Gene/Gene ID	Sequence (5'-3')	Tm° (°C)
PBB2	F: TGCCGAAGAAACGCAATGCTTCAA	74.7
Glyma18g51320	R: TGCAGCAAGTGAACCTGATCCCAT	72.6
COL2a	F: TGGTTCTGTTAGTCAAAGTGTTTCAG	64.2
Glyma08g28370	R: GCATCCGATTAGTACAACTGCAAG	65.5
COL2b	F: TTCCTCTGCAAGGCTGACGC	70.1
Glyma08g28370	R: GTGCTGATGAGAGTAATGGTCCGT	67.3
COL9	F: GTCTCCGAACTCTTGTCCATC	62.7
Glyma19g05170	R: CACACGAACATATCCGAAAGC	64.3
COLX	F: GCTGAAGTGGACCAGATGTT	61.8
Glyma13g07030	R: TGCTTCCAATGATGTGGTTC	63.6
COL5	F: AATTGACTCGGATGTGGAGCGTCT	70.8
Glyma13g01290	R: TGGCTTGGTGTGACTTTGGAATGC	72.8
COLY	F: CATGCTGGACACTCCATACG	64.2
Glyma17g07420	R: GCGCTCCAGCTAAGAATCAC	63.9
COL4	F: TCAGATGAGCCAGAGTGTATCGTC	66.6
Glyma04g06240	R: AGCGAAATCAACACGACGGAACGA	74.1
COLZ	F: ACGCATCGAGAAAAGCGTAT	63.6
Glyma06g06300	R: ACAGCAGCAAAAGCAAACAT	62.4
COL92	F: TGTCCCACTCAAAGTCACCA	64.4
Glyma19g39460	R: CTCTGCATAGGCCTTCCTTG	63.8

Table 2. List of primers with gene names and gene IDs (phytozome.net), sequences (5'-3'): forward (F) and reverse (R) primers, and melting temperature (Tm° (°C)).

Gene	Abbrev.	Gene ID	Annotation
Arabidopsis thaliana			
CONSTANS	*CO*	At5g15840	B-box type zinc finger protein with CCT domain
CONSTANS-like *1*	*AtCOL1*	At5g15850	B-box type zinc finger protein with CCT domain
CONSTANS-like *2*	*AtCOL2*	At3g02380	B-box type zinc finger protein with CCT domain
CONSTANS-like *3*	*AtCOL3*	At2g24790	B-box type zinc finger protein with CCT domain
CONSTANS-like *4*	*AtCOL4*	At5g24930	B-box type zinc finger protein with CCT domain
CONSTANS-like *5*	*AtCOL5*	At5g57660	B-box type zinc finger protein with CCT domain
CONSTANS-like *6*	*AtCOL6*	At1g68520	B-box type zinc finger protein with CCT domain
CONSTANS-like *7*	*AtCOL7*	At1g73870	B-box type zinc finger protein with CCT domain
CONSTANS-like *8*	*AtCOL8*	At1g49130	B-box type zinc finger protein with CCT domain
CONSTANS-like *9*	*AtCOL9*	At3g07650	B-box type zinc finger protein with CCT domain
CONSTANS-like *10*	*AtCOL10*	At5g48250	B-box type zinc finger protein with CCT domain
CONSTANS-like *11*	*AtCOL11*	At4g15250	B-box type zinc finger protein with CCT domain
CONSTANS-like *12*	*AtCOL12*	At3g21880	B-box type zinc finger protein with CCT domain
CONSTANS-like *13*	*AtCOL13*	At2g47890	B-box type zinc finger protein with CCT domain
CONSTANS-like *14*	*AtCOL14*	At2g33500	B-box type zinc finger protein with CCT domain
CONSTANS-like *15*	*AtCOL15*	At1g28050	B-box type zinc finger protein with CCT domain
CONSTANS-like *16*	*AtCOL16*	At1g25440	B-box type zinc finger protein with CCT domain
Oryza sativa			
Heading date 1	*Hd1*	Os06g16370	CCT/B-box zinc finger protein, expressed
Glycine max			
CONSTANS-like *2a*	*COL2a*	Glyma18g51320	B-box type zinc finger protein with CCT domain
CONSTANS-like *2b*	*COL2b*	Glyma08g28370	B-box type zinc finger protein with CCT domain
CONSTANS-like *9*	*COL9*	Glyma19g05170	B-box type zinc finger protein with CCT domain
CONSTANS-like *X*	*COLX*	Glyma13g07030	B-box type zinc finger protein with CCT domain
CONSTANS-like *5*	*COL5*	Glyma13g01290	B-box type zinc finger protein with CCT domain
CONSTANS-like *Y*	*COLY*	Glyma17g07420	B-box type zinc finger protein with CCT domain
CONSTANS-like *4*	*COL4*	Glyma04g06240	B-box type zinc finger protein with CCT domain
CONSTANS-like *Z*	*COLZ*	Glyma06g06300	B-box type zinc finger protein with CCT domain
CONSTANS-like *92*	*COL92*	Glyma19g39460	B-box type zinc finger protein with CCT domain

Table 3. Genes used in Phylogenetic Analysis.

AT5G15840	RPCDTCRNACTVYCADAYLCCDVHSARHRVCPAFNPRPGTVREARVRYREKRTRFEKTIRYSRKYAERRVNGRFAKR
AT3G02380	RACDTCRAACTVYCADAYLCCDVHAARHRVCPAFNPRPAIVREARVRYREKKTRFDKTIRYSRKYAERRIKGRFAKR
AT5G15850	QACDTCRAACTVYCADAYLCCDVHAARHRVCPAFNPRPELVREARVRYREKKMRFEKTIRYSRKYAERRIKGRFAKK
Glyma13g07030	HVCDTCRAPCVLYCADAYLCCDVHAARHRVCPAFNPRPGVVREARVRYREKKTRFEKKIRYSRKYAERRIKGRFAKR
Glyma19g05170	RVCDTCLAPCVLYCADAYLCCDVHAARHRVCPAFNPRPGVVREARVRYREKKTRFEKKIRYSRKYAERRIKGRFAKR
Os06g16370	RPCDGCRAPSVVYCADAYLCCDVHAARHRVCPALNPTPAAIREARVRYREKKARFEKTIRYTRKYAERRIKGRFAKR
Glyma08g28370	RMCDTCRAPSSVFCAHAFLCCDLHALWHRVCPAFNPRPAIVREARVRYREKKTRFEKTIRYSRKYAERRIKGRFAKR
Glyma18g51320	RMCDTCRVPSTVFCSHAFLCCDLHVLWHRVCPAFNPRPAIVREARVRYREKKMRFEKTIRYSRKYAERRIKGRFAKR
AT5G24930.1	KLCDSCKATAALYCPDAFLCCDVHAARHRVCPAVNPRPDVVREARVRYREKRNRFEKTIRYSRKYAERRIKGRFAKR
Glyma04g06240	KLCDSCKATATLYCPDAFLCCDVHAARHRVCPAVNPRPEVVREARVRYREKRNRFEKTIRYSRKYAERRIKGRFAKR
Glyma06g06300	KLCDSCKATATLYCPDAFLCCDVHAARHRVCPAVNPRPEVVREARVRYREKRNRFEKTIRYSRKYAERRIKGRFAKR
Glyma13g01290	KPCDSCKASAALFCLDAFLCCDIHCARHRVCPAVNPRPDVVREARVRYREKRNRFEKTIRYSRKYAERRIKGRFAKR
Glyma17g07420	KPCDSCKASAALFCPDAFLCCDIHCARHRVCPAVNPRPDVVREARVRYREKRNRFEKTIRYSRKYAERRIKGRFAKR
AT2G24790	RLCDSCKTAATLFCADAFLCCDIHTARHRVCPAVNPRPDVLREARVRYREKRNRFEKTIRYSRKYAERRIKGRFAKR
AT5G57660	RSCDACKVTAAVFCVDAFLCCDIHSFRHRVCPAVNPRPDYVREARVRYREKRNRFEKTIRYSRKYAERRIKGRFAKR
Cre13.g596300	SSCVVCAAAAVVWCNDALLCCDIHTARHRFCGALNPTPSFAREQRVRYREKRNRFAKTIRYSRKYAERRIKGRFAKK
AT1G68520	RACDSCVRRARWYCADAFLCCDVHSARHRVKAGYTPLPDYSREARVRYREKRTRFSKKIRYVRKNAERRMKGRFVKR
AT1G25440	RACDSCVRRARWYCADAFLCCDVHSARHRVKPAVTPLPDYSREARVRYREKRTRFSKKIRYVRKNAERRMKGRFVKR
AT1G73870	RGCDACMSRASWYCADAFLCCDIHSARHRVQPTTTPKLNYAREARVRYKEKRTRFSKKIRYVRKNAERRIKGRFVKR
AT1G49130	RACELCLKHAVWYCSDAFLCCDVHSAKHRVCSNVTPKPDYNREARVRYRDKRNRFEKKIRYVRKNADRRMKGRFVRR
AT2G47890	RLCDYCDSVALVYCADAKLCCDVHVFKHRSCPSLTARPGGLRNSALRYKEKKSRYEKHIRYSRKRAERRIRGRFAKA
Glyma19g39460	RPCDYCGSTALLYCADAKLCCDVHSFKHRTCPAIALRPGLTRDSALRYKQKKTRFDKHIRYSRKRAERRVKGRFAKM
AT1G28050	VPCDFCGRTAVLFCADAKLCCDVHTSKHRSCPVVCSRAGGVRGDAMRYKEKRTRYDKTIRYSRKRADRRVRGRFVKA
AT2G33500	VACEFCGRTAVLFCADAKLCCDVHSSKHRSCPVVCSRAGTGRDNAMRYKEKKTRYDKTIRYTRKRAERRVKGRFVKA
AT5G48250	YMCDFCGQRSMVYCSDACLCCDVHSSRHRTCPAVSTRTGYSRNNAVRYKEKKARFDKRVRYSRKRADRRVKGRFVKS
AT3G07650	YMCDFCGQRSMVYCSDACLCCDVHSSRHRTCPAVSTRTGFSRNNAVRYKEKKARFDKRVRYSRKRADRRVKGRFVKA
AT4G15250	ARCDFCGEKALIYCSDAKLCCDVHSSRHRSCPTVNCLSDGVRDEAKRYKQKKKRFGKQIRYSRKRADRRVKGRFVKS
AT3G21880	PKCDHCASQALIYCSDAKLCCDVHSSRHRSCPAINCVSGGIRNEAKRYKEKKKRFGKQIRYSRKRADRRVKGRFVKA

Table 4. Manually edited multiple sequence alignment of amino acid sequences of proteins from Table 1 used for phylogenetic analysis in Figure 9.

		Line	Clark		Williams82		L65-3366		L66-432		L74-441		L92-1195		*Glycine soja*	
			$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$
COL2a	SD	6:30	787.89	16.06	818.5	27.246	712.785	45.981	760.255	5.76	875.55	23.30	714.78	25.89	950.7	19.0
		14:30	29.35	2.04	29.425	2.682	33.925	0.763	31.020	1.927	29.315	0.102	31.415	0.510	26.08	1.41
		22:30	7.87	0.60	12.43	0.66	12.79	1.02	11.45	0.48	12.34	0.53	14.02	0.53	11.37	0.20
	LD	6:30	146.77	5.28	150.89	5.73	143.01	9.87	154.88	7.28	163.42	10.35	165.81	9.08	141.98	5.31
		14:30	10.99	1.13	11.91	0.26	12.91	0.43	17.70	0.22	10.27	0.37	14.31	1.60	17.50	0.36
		22:30	245.65	13.73	179.66	10.21	202.50	4.63	165.53	8.46	255.02	2.89	206.94	0.66	154.30	1.86
COL2b	SD	6:30	330.32	28.48	301.72	1.92	283.39	3.95	306.82	10.79	335.71	2.82	314.80	2.75	334.70	16.1
		14:30	15.92	0.004	12.02	0.89	17.46	0.01	14.63	0.24	14.92	0.30	17.95	0.11	9.98	0.39
		22:30	5.45	0.09	7.36	0.18	7.77	0.02	8.85	1.24	7.42	0.12	7.61	0.09	5.69	0.34
	LD	6:30	59.76	3.57	64.24	8.89	65.86	4.86	58.55	1.76	50.10	3.22	77.80	4.14	63.06	2.39
		14:30	10.29	0.66	9.47	0.33	11.35	0.58	9.93	1.56	7.71	0.15	14.64	0.73	9.89	0.17
		22:30	113.16	8.15	86.09	2.87	75.43	2.72	93.09	4.14	104.45	3.35	80.51	0.99	68.33	1.02
COL9	SD	6:30	0.44	0.02	0.63	0.00	0.37	0.04	0.31	0.06	0.12	0.12	0.29	0.01	0.19	0.07
		14:30	0.12	0.02	0.18	0.06	0.04	0.04	0.13	0.01	0.14	0.04	0.19	0.02	0.06	0.03
		22:30	0.49	0.02	0.55	0.02	0.48	0.07	0.53	0.003	0.60	0.21	0.34	0.08	0.54	0.23
	LD	6:30	0.97	0.01	0.84	0.07	0.96	0.31	0.91	0.37	0.93	0.07	1.57	0.27	0.56	0.03
		14:30	2.45	0.14	2.45	0.27	1.75	0.13	1.12	0.05	1.40	0.07	3.51	0.17	3.15	0.34
		22:30	0.86	0.07	0.67	0.11	0.59	0.07	0.27	0.17	0.40	0.12	0.18	0.06	0.24	0.004
COLX	SD	6:30	0.38	0.14	1.02	0.05	0.38	0.04	1.08	0.11	0.77	0.04	1.33	0.04	0.41	0.004
		14:30	0.22	0.03	0.17	0.04	0.05	0.04	0.18	0.05	0.05	0.04	0.12	0.10	0.07	0.04
		22:30	0.93	0.04	1.01	0.02	1.35	0.22	0.95	0.16	0.90	0.02	0.65	0.16	1.04	0.11
	LD	6:30	0.39	0.09	0.19	0.06	0.46	0.08	0.44	0.06	0.40	0.01	0.87	0.08	0.20	0.04
		14:30	2.00	0.16	1.55	0.20	1.17	0.06	0.42	0.02	0.77	0.04	1.64	0.03	1.34	0.08
		22:30	0.54	0.09	0.27	0.18	0.61	0.03	0.35	0.05	1.09	0.06	0.22	0.02	0.40	0.05

Table 5. RNA-Seq results for all nine CO-like homologs (*COL2a*, *COL2b*, *COL9*, *COLX*, *COL5*, *COLY*, *COL4*, *COLZ*, and *COL92*) in seven different soybean genotypic backgrounds (Table 1) sampled from three time points (6:30, 14:30, and 22:30) for two different day length treatments (SD and LD). "$\bar{x}$" represents the mean of the RPKMs for three biological replications. *"$s_{\bar{x}}$"* represents the standard error of the mean "$\bar{x}$" using three biological replications.

		Line	Clark		Williams82		L65-3366		L66-432		L74-441		L92-1195		*Glycine soja*	
			$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$
COL5	SD	6:30	130.17	14.89	105.34	5.38	120.69	5.60	150.41	6.92	141.33	5.83	132.04	1.91	135.57	7.64
		14:30	180.21	4.69	109.96	7.21	167.16	15.30	165.26	3.96	174.97	3.08	151.62	0.45	157.21	3.61
		22:30	12.59	0.27	9.41	0.78	19.90	0.71	16.80	4.55	13.30	0.37	8.27	0.18	15.79	3.40
	LD	6:30	135.17	5.71	152.06	18.64	202.27	0.12	131.41	1.55	187.22	16.56	124.36	15.86	198.79	1.40
		14:30	38.67	2.58	33.70	3.72	33.29	0.73	33.61	0.65	34.97	3.45	31.79	1.04	39.17	1.11
		22:30	23.70	2.69	23.66	1.16	22.80	0.36	32.37	3.85	23.71	0.66	21.32	2.17	19.65	0.96
COLY	SD	6:30	84.53	2.87	60.65	2.74	62.88	6.65	87.55	5.44	70.55	0.62	65.15	1.12	61.23	0.51
		14:30	136.93	3.66	152.07	6.14	139.02	9.42	149.20	0.41	137.46	4.30	139.13	2.19	131.16	2.04
		22:30	2.67	0.05	2.88	0.24	4.79	0.13	6.30	0.62	3.33	0.18	2.56	0.26	6.03	0.13
	LD	6:30	83.08	0.28	83.41	6.79	100.83	5.70	82.58	3.90	108.38	9.77	99.68	7.64	119.34	2.18
		14:30	18.93	0.77	20.80	1.48	23.18	0.09	21.53	1.72	14.43	1.82	17.45	0.65	35.38	2.27
		22:30	9.86	0.19	9.17	0.02	7.73	0.07	8.28	0.13	9.50	0.26	7.80	0.26	8.29	0.20
COL4	SD	6:30	167.46	15.81	171.43	4.70	132.15	9.52	159.55	4.43	185.50	4.54	155.19	1.68	157.81	3.31
		14:30	36.41	1.81	33.38	0.19	36.03	0.15	35.05	0.89	35.64	0.19	34.53	2.30	37.89	2.06
		22:30	29.23	0.93	28.41	0.13	30.49	1.92	29.34	0.64	26.01	1.88	27.26	0.50	42.52	1.00
	LD	6:30	87.03	2.72	96.11	4.31	90.63	2.85	82.42	0.29	92.74	2.04	99.97	3.97	72.79	2.59
		14:30	23.35	1.72	29.47	1.20	19.02	1.78	24.49	1.47	19.61	0.83	22.76	0.71	33.08	0.85
		22:30	96.57	4.48	76.51	2.41	82.62	2.38	70.70	0.07	107.85	2.71	75.69	0.62	68.10	1.69
COLZ	SD	6:30	311.48	12.09	336.36	18.56	285.51	5.04	321.03	7.92	414.44	28.80	337.70	5.80	389.96	22.77
		14:30	71.87	1.77	71.64	1.43	72.58	0.15	70.44	2.96	78.92	0.10	74.38	3.89	87.60	0.53
		22:30	72.70	1.09	63.20	2.31	74.75	3.29	68.72	2.90	78.17	4.53	65.41	1.37	107.03	2.77
	LD	6:30	155.22	2.38	175.35	13.00	160.26	7.54	151.18	2.31	165.43	1.77	181.11	5.09	132.54	3.21
		14:30	63.57	0.34	66.44	3.69	64.77	3.59	79.80	1.10	53.49	4.13	73.84	0.73	79.48	6.67
		22:30	162.61	7.72	164.77	1.14	164.37	10.05	185.68	10.39	240.37	1.67	186.26	13.07	162.39	7.34
COL92	SD	6:30	6.85	0.47	7.61	0.81	7.05	0.81	7.20	0.14	8.51	0.25	6.70	0.41	6.13	0.59
		14:30	9.83	0.15	9.18	0.004	10.70	0.59	9.53	0.29	8.99	0.05	8.41	0.35	8.28	0.13
		22:30	12.69	0.66	12.04	0.41	15.87	0.53	11.85	0.42	14.70	0.44	14.01	0.10	11.25	0.53
	LD	6:30	12.44	1.34	12.75	0.18	7.31	0.50	10.59	0.31	7.03	0.18	14.22	0.35	5.16	0.28
		14:30	10.80	0.04	9.87	0.19	11.32	0.15	14.75	1.14	8.32	1.26	12.56	0.89	10.50	0.29
		22:30	10.76	0.33	9.41	0.19	9.21	0.31	9.72	0.16	11.92	0.49	8.89	0.64	9.70	0.06

Table 5. continued.

		Line	Clark		Williams82		L65-3366		L66-432		L74-441		L92-1195		*Glycine soja*	
			$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$
GmFT2a	SD	6:30	40.45	1.158	36.15	2.357	35.47	0.782	25.04	0.672	29.07	0.285	33.62	4.380	26.80	4.707
		14:30	59.07	4.894	41.50	2.399	42.45	0.990	43.91	2.944	42.69	2.206	41.00	1.314	35.67	2.378
		22:30	16.77	2.586	22.52	1.882	15.92	0.847	11.20	2.665	26.06	1.338	21.42	0.285	19.98	2.512
	LD	6:30	0.261	0.136	0.327	0.234	0.053	0.053	0.239	0.148	0.261	0.070	0.000	0.000	0.421	0.055
		14:30	0.096	0.096	0.062	0.062	0.061	0.061	0.136	0.068	0.277	0.226	0.176	0.176	0.091	0.091
		22:30	0.457	0.329	0.083	0.083	0.138	0.069	0.066	0.066	0.305	0.069	0.000	0.000	0.058	0.058
GmFT5a	SD	6:30	85.86	1.993	66.71	2.079	59.83	3.219	72.76	2.985	69.74	6.753	65.07	3.184	43.14	0.794
		14:30	23.92	2.590	29.36	1.420	22.54	2.986	24.39	0.712	21.64	1.348	21.71	1.737	21.25	2.169
		22:30	15.36	1.191	20.03	2.075	19.64	3.195	12.07	1.050	19.32	1.201	13.82	0.545	15.71	0.644
	LD	6:30	1.199	0.893	0.840	0.591	0.078	0.064	0.668	0.206	0.965	0.202	1.405	0.540	0.074	0.074
		14:30	0.658	0.160	0.829	0.140	0.183	0.183	0.278	0.176	0.828	0.114	0.583	0.198	0.000	0.000
		22:30	0.910	0.482	1.120	0.274	0.288	0.068	0.273	0.083	3.677	0.685	0.369	0.119	0.057	0.057

Table 6. RNA-Seq results for the two *FT-like* homologs (*GmFT2a* and *GmFT5a*) in seven different genotypic backgrounds (Table 1) sampled from three time points (6:30, 14:30, and 22:30) for two different day length treatments (SD and LD). "$\bar{x}$" represents the mean of the RPKMs for three biological replications. "$s_{\bar{x}}$" represents the standard error of the mean "$\bar{x}$" using three biological replications.